On the Computational Power of Simple Dynamics

Emanuele Natale

SAPIENZA
UNIVERSITÀ DI ROMA

Sapienza, Università di Roma

Dottorato di Ricerca in Computer Science

XXIX Ciclo - 2016

On the Computational Power of Simple Dynamics

Emanuele Natale

Date of Thesis Defense:
13 February 2017

SAPIENZA
UNIVERSITÀ DI ROMA

Sapienza, Università di Roma

Dottorato di Ricerca in Computer Science

XXIX Ciclo - 2016

Emanuele Natale

On the Computational Power of Simple Dynamics

Thesis Committee

Prof. Andrea Clementi (Advisor)
Prof. Riccardo Silvestri (Advisor)
Prof. Flavio Chierichetti
Prof. Pierluigi Crescenzi

To my family

Abstract

This work presents a set of analytical results regarding some elementary randomized protocols, called *dynamics*, for solving some fundamental computational problems. New techniques for analyzing the processes that arise from such dynamics are presented, together with concrete examples on how to build efficient and robust distributed algorithms for some important tasks using these processes as a black-box.

More specifically, we analyze several dynamics such as the 3-Majority, the Averaging and the Undecided-State ones, and we show how to use them to solve fundamental problems such as plurality consensus, community detection (including the reconstruction problem in the stochastic block model), and bit dissemination (rumor spreading). We focus mainly on unstructured and random interaction models, and we also deal with scenarios in which the communication is affected by noise or when a self-stabilizing protocol is required.

Preface

This work presents in a systematic way a major part of the research I've taken part to during my PhD studies. The main purpose of this Preface is to list what has been included here out of what I've done in these three years, and what has been not because of the diversity of topic.

A great part of such work has already been presented at conferences in computer science. The following chapters are based on the following publications:

- Chapter 4: L. Becchetti, A. Clementi, E. Natale, F. Pasquale, and L. Trevisan, *Find Your Place: Simple Distributed Algorithms for Community Detection,* in Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (*SODA'17*), Barcelona, Spain, 2017.
- Chapter 5:
  - L. Becchetti, A. Clementi, E. Natale, F. Pasquale, R. Silvestri, and L. Trevisan, *Simple dynamics for plurality consensus,* in Proceedings of the 26th ACM on Symposium on Parallelism in Algorithms and Architectures (*SPAA'14*), Prague, Czech Republic, 2014, pp. 247–256.
  - L. Becchetti, A. Clementi, E. Natale, F. Pasquale, and L. Trevisan, *Stabilizing Consensus with Many Opinions,* in Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (*SODA'16*), Arlington, Virginia, 2016, pp. 620–635.
- Chapter 6: L. Becchetti, A. Clementi, E. Natale, F. Pasquale, and R. Silvestri, *Plurality Consensus in the Gossip Model,* in Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (*SODA'15*), San Diego, California, 2015, pp. 371–390.
- Chapter 7: L. Becchetti, A. Clementi, E. Natale, F. Pasquale, and G. Posta, *Self-Stabilizing Repeated Balls-into-Bins,* in Proceedings of the 27th ACM on Symposium on Parallelism in Algorithms and Architectures (*SPAA'15*), Portland, Oregon, 2015, pp. 332–339.
- Chapter 8: P. Fraigniaud and E. Natale, *Noisy Rumor Spreading and Plurality Consensus,* in Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing (*PODC'16*), Chicago, Illinois, 2016, pp. 127–136.

- Chapter 9: L. Boczkowski, A. Korman, and E. Natale, *Minimizing Message Size in Stochastic Communication Patterns: Fast Self-Stabilizing Protocols with 3 bits,* in Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (*SODA'17*), Barcelona, Spain, 2017. (Brief Announcement in Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing (*PODC'16*). Chicago, Illinois, 2016, pp. 207–209.)

As one can see in the previous list, rather than following the chronological order of the research, the presentation of the material attains to the "big picture" discussed in the Introduction (Chapter 1).

A paper which *could* have been included in this work is

- D. Kaaser, F. Mallmann-Trenn, and E. Natale, *On the Voting Time of the Deterministic Majority Process*, in Proceedings of the 41st International Symposium on Mathematical Foundations of Computer Science, Dagstuhl, Germany, 2016, vol. 58, p. 55:1–55:15. (MFCS'16)

While the subject of the aforementioned paper is akin to the processes investigated here, the deterministic and worst-case nature of those results does not fit in the spirit of this work, as explained in the Introduction (Chapter 1).

Two other papers which don't have anything to do with distributed computing are

- L. Guala, S. Leucci, and E. Natale, *Bejeweled, Candy Crush and other match-three games are (NP-)hard*, in Proceedings of the IEEE Conference on Computational Intelligence and Games, 2014, pp. 1–8. (CIG'14)
- L. Gualà, S. Leucci, E. Natale, and R. Tauraso, *Large Peg-Army Maneuvers*, in Proceedings of the 8th International Conference on Fun with Algorithms, Dagstuhl, Germany, 2016, vol. 49, p. 18:1–18:15. (FUN'16)

The previous papers deal with computational aspects of some combinatorial puzzles, and they came out of the common interest of Luciano Gualà, Stefano Leucci, Roberto Tauraso and me for algorithmic aspects of perfect-information single-player games.

Last but not least, the following work by Michele Borassi and me on computing the betweenness centrality of complex networks is also off-topic with respect to the scope of this treatise:

- M. Borassi and E. Natale, *KADABRA is an ADaptive Algorithm for Betweenness via Random Approximation*, in Proceedings of the 24th Annual European Symposium on Algorithms, Dagstuhl, Germany, 2016, vol. 57, p. 20:1–20:18. (ESA'16)

Acknowledgments

Contrary to the old saying that "the neighbor's grass is always greener", every time I look at any neighbor's grass I feel so lucky for having had such an amazing garden.

There is no way I can express my gratitude to Andrea Clementi for having been an advisor so dedicated to my best scientific growth.

A special thanks goes to my internal advisor Riccardo Silvestri: being unconditionally exposed to his point of views has saved me from taking many erroneous decisions.

I thank my "academic older brother" Francesco Pasquale and Luca Becchetti for the great time during our research meetings.

I thank Luca Trevisan for the invaluable opportunities he has given me, the first of which is that of working with him.

I thank some professors which enriched my love for mathematics and computer science and with whom later I also had the pleasure to do research with, such as Roberto Tauraso, Miriam Di Ianni, Giorgio Gambosi and Luciano Gualà.

I thank Robert Elsässer and Petra Berenbrink for the great time in Salzburg and Hamburg, respectively.

I thank Pierre Fraigniaud for the invaluable time I had in Paris, where I also met Amos Korman. Working with them has been a very rewarding experience on several levels.

These years would have not been the same without other PhD students sharing with me an important part of them, such as Stefano Leucci, Dominik Kaaser, Frederik Mallmann-Trenn, Lucas Boczkowski and Michele Borassi. Thanks to all of you.

I thank Alessandro Panconesi and Flavio Chierichetti for supporting a great part of the research that is presented in this work.

I thank my partner for having walked by my side along this long journey, supporting me in the difficult moments. A special thanks also to her family.

I deeply thank my loving family for having done all their best to allow me to do what I wanted to do.

Contents

List of Figures

List of Definitions, Theorems and Corollaries

CHAPTER 1

Introduction

This work is a treatise in the field of distributed computing and, as such, there are some expectations that we are not going to disappoint: there is a system (the network) of n agents (the nodes) that interact (exchange messages) with each other according to some communication model, and there is a computational goal that the system aims to achieve through some suitable *protocol* executed by the agents [**Pel00**].

Within the field of distributed computing, the scope of this work is located within a class of systems that may resemble the subject of study of statistical mechanics [**Lig12**]: the class of protocols that we consider are *simple* and *lightweight* [**HP01**], their typical behavior strongly relies on randomness which constitutes an essential part of the process, and have been grouped under the name of *dynamics* [**AAE08, AAB$^+$11, Dot14, MNT14**].

As in the case of *natural algorithms* and *complex networks*, the concept of dynamics seems affected by a clear contrast between the informal consensus the related experts' community has about the obviousness of what that concept means, and the lack of serious attempts to provide a rigorous definition which can englighten the outsiders.

To prevent us from contributing to such undesirable situation, with the first definition of this work we attempt to provide a first formalization[1] of the concept of *dynamics* as *simple, lightweight, natural, local, elementary rules.*

Definition 1 (Dynamics). A *dynamics* is a synchronous distributed algorithm characterized by a very simple structure, whereby the state of a node at round t depends only on her state and a symmetric function of the multiset of states of her neighbors at round $t-1$, while the update rule is the same for every graph and every node and does not change over time.

Remark 1. Clearly, within the constraints of the previous definition, it still appears to be possible to come up with computational rules which are *complex and unnatural.* We emphasize that the nature of Definition 1 is to provide a rough guideline, and does not substitute the reliance of the scientific community on the real world phenomena the concept intends to capture, which are discussed in Chapter 3. Definition 1 is therefore overtly provisional and open to be replaced by better candidates.

[1] The definition has already appeared in [**BCN$^+$15b**].

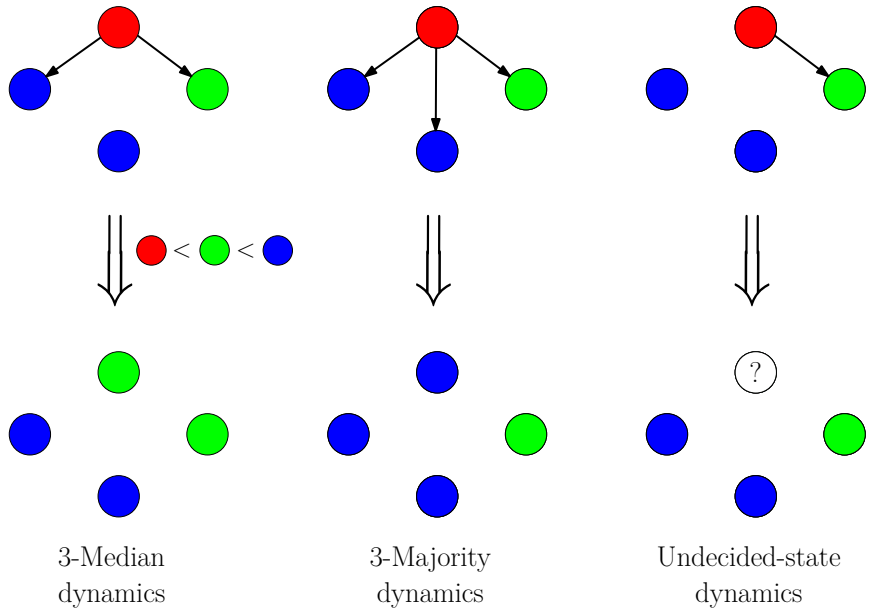

FIGURE 1. Illustration of the 3-Median dynamics (in which each agent samples two other agents at random and updates her opinion with the median of their values and her own), the 3-Majority (in which each agent samples three other agents at random and update her opinion with the most frequent value among those three, breaking ties arbitrarily), and Undecided-State Dynamics (in which each agent samples another agent, if their values differ she becomes *undecided*, and if she is undecided she picks the first opinion she sees).

Note that in Definition 1 no network IDs are used, so we may assume that the network is *anonymous*. Examples of dynamics which are discussed in Chapter 3 include update rules in which every node updates its state to the plurality or the median of the states of its neighbors[2], or which updates it to the average of the values held by its neighbors (see Figure 1). In contrast, an algorithm that, say, proceeds in two phases, using averaging during the first $10 \log n$ rounds and plurality from round $1 + 10 \log n$ onward, with n the number of nodes, is not a dynamics according to our definition, since its update rule depends on the size of the graph. As another example, an algorithm that starts by having the lexicographically first node elected as "leader" and then propagates her state to all other nodes again does not meet the definition of dynamics, since it assigns roles to the nodes and requires them to possess distinguishable identities.

[2]When states correspond to rational values.

Organization of the work

In Section 1.1, we begin our journey with an informal account of the content of this work. In the following chapter (sections 2.1, 2.2, 2.3, 2.4, 2.5 and 2.6), we present the results contained in this work. In Chapter 3, we discuss the related literature. In chapters 4, 5, 6, 7, 8 and 9, we present the proofs of our results. Finally, in Chapter 10, we discuss some open problems.

1.1. The Informal Story of the Big Picture

In sections 2.1, 2.2, 2.3, 2.4, 2.5 and 2.6, we are going to individually motivate the subject of each of the following chapters. However, as we said above, all of them can be framed within the investigation of the computational power of dynamics. Hence, apart from the individual motivations, the question of whether it makes sense to look at them as a coherent whole arises naturally. Therefore, before providing further details on the individual subjects of our study and on the empirical reasons that motivate our specific interest for them, it is worth taking a small digression about how we ended up looking at them as belonging to a consistent class of objects, where each of them is deeply intertwined to the others.

1.1.1. Peeking in the Universe of Computational Rules

Since the advent of the computer, scientists found themselves with a new telescope which provided them with the capacity to observe the computational universe at a new scale. Through simulations, they were able to look far beyond their mathematical understanding of the relationship between local interactions among the tiny parts of a system and its global behavior and, in the last decades, they were astounded by the unexpected appearance of global complexity from local simplicity. To mention few examples of the enthusiasm of the scientific community, in 1984 the Santa Fe Institute was founded in New Mexico, with the mission of pioneering research on how complex systems emerge from simple interactions and, almost twenty years later, Stephen Wolfram was publishing his famous book [**Wol02**], in which he provided extensive empirical evidence about the fact that many complex systems *emerge* from very simple "programs". However, such enthusiasm brought from the shocking new ability to explore the universe of computational rules has been counterbalanced from the inability to develop a new mathematics which could account for our new observations.

With the above perspective in mind it is hard not to be fascinated by the difficulty of saying something mathematically nontrivial on the "complexity from simplicity" phenomenon (in short CFS phenomenon). One of the few possible paths in the latter direction with a non-negligible probability of being profitable, appears to be that of theoretical computer science. The mathematics of computation, which made us concretely aware of the CFS problem, seems one of the few sensible theoretical tools on which to bet for understanding it.

Within the world of theoretical computer science, a particularly appealing tool to look at the interplay of local/individual and global/collective behavior is the *theory of distributed computing* (in short, distributed computing). Distributed computing is concerned with how systems of computational agents can achieve some global goal in the most efficient way. If we set as the goal of the system "a complex behaviour" and we constrain the agents to perform only "simple" interactions and computations, we get an instance of the CFS phenomenon. Therefore, in some sense, we have an entire subfield of theoretical computer science (and thus, of mathematics), whose purpose is (in part) to explicitly deal with the CFS phenomenon.

The above interpretation of the status quo is not wishful thinking. From programmable matter [**DDG+14, CDRR16**] to chemical reaction networks [**CSWB09, Dot14, CKW16, Reu16**], from sensor networks [**AAD+06, AFJ06**] to the behaviour of insect colonies [**FHK14, FN16**], there is a huge part of the distributed computing discipline driven by the aspiration to develop a theory analogous to that built by statistical mechanics for interacting particle systems, when we replace "particle" with "agent".

In fact, the underlying motif behind the research presented in this work arose when the author said to Andrea Clementi (who was teaching a course on distributed computing that the author was attending), that he would have liked to work on a problem which consisted in finding a simple process whose interest lied in the intersection of distributed computing and network analysis, i.e. that would have shown some complex behaviour depending on the network topology. Andrea Clementi came up with a problem that, as we later discovered, turned out to be an instance of the famous reconstruction problem in stochastic block models [**HLL83, DF89, JS98, McS01, CO10, DKMZ11, ABH14**], which is a main character of Chapter 4. In the next section we informally discuss the original problem and how the different results of this work can be traced back to the first natural idea with which we tried to solve it.

1.1.2. Dynamics for Distributed Clustering and Much More

Consider the problem of performing community detection on a model of (discrete-time) dynamic random graphs [**AKL08**], the *dynamic stochastic block model*, which is obtained by considering a sequence of independent graphs generated according to a fixed stochastic block model (see Definition 6). That is, in the dynamic stochastic block model the nodes are partitioned in two *communities* of equal size and at each round a random graph is generated by including each edge between nodes within the same community with probability p, and each edge across the two communities with probability $q < p$. It follows that each node tends to have more neighbors inside her own community than the other one.

To perform community detection means to assign to each node a label such that two nodes have the same label if and only if they are in the same

community[3]. Note that the previous definition, in general, still requires to define what a community is. A natural way to address this issue is to consider a *planted* model, i.e. to include the communities right in the definition of the graph model, as is the case of the aforementioned dynamic stochastic block model and the graph models considered in 4.

In such a scenario a natural heuristic that comes to one's mind to solve the problem is the following:

(1) Each node initially generates a random color;
(2) At each round each node takes the most frequent value of a random sample of neighbors, chosen independently and uniformly at random, breaking ties arbitrarily.

The previous family of *epidemic* strategies and their variants are known as *label propagation algorithms* (LPA for short) [**RAK07, BC09, LHLC09, LM10**]. The intuition is that the mechanism employed in the second part of the algorithm[4] should tend to assign the same color to sets of nodes which are more connected among themselves than with the rest of the graph.

Perhaps surprisingly, the rules in Step 2 which (experimentally) turns out to be the most effective, efficient and robust are probabilistic rules that, in an infinite time, would lead the system to a trivial labeling. In other words, there is a possible (although exponentially improbable) concatenation of specific unlikely events in the random choices of the protocol which could lead the system to a complete failure, such as labeling the whole graph as a sole community. This scenario is often encountered also in other scientific contexts such as systems studied in statistical mechanics where, a priori, an "almost-impossible" sequence of unfortunate events would cause an empirical violation of the laws of thermodynamics. To cope with such *bad* events the concept of *metastability* has therefore been introduced. A set of states of a stochastic process is said to be metastable if, informally speaking, the system spends *a lot of time* in that class of states, although they may be far from those that the system reaches in the *equilibrium*, i.e. in an infinite time (in the language of Markov chains, the metastable states may even be transient, i.e. once the system exits them, it never visits them again). Analogously, a good LPA is expected to assign (with high probability) the same label to nodes in the same community, and different labels to nodes in different communities, and to maintain this status of *internal* consensus and *external* disagreement for any polynomial number of rounds although, in an exponential time, it may be that the system happens to assign the same label to different communities, with no possibility of recovering from that point on. Thus, the efficacy of LPAs in solving the community detection problem partly relies on the efficacy and robustness of the employed mechanism in

[3]According to the literature discussed in Chapter 3, in this work we assume that the communities partition the graph, i.e. each node belongs to exactly one community.

[4]Typically, the update rule of an LPA make use of is a dynamics.

cautiously solving the plurality consensus problem, that is the problem of converging to the most frequent color in the system[5] (see Section 2.2).

Despite their extreme simplicity, the analysis of LPA-based protocols is extremely challenging, as discussed in Section 3.1.1. In fact, not surprisingly, while simulations were decisively promising for simple variants of the previous protocol (such as when in Step 2 we adopt the 3-Majority dynamics discussed in Chapter 5), in [**CDIG$^+$15**] we manage to rigorously analyze only a distributed community detection algorithm which is quite far from being a dynamics, given that the rule it applies changes as a function of time which depends on the number of nodes n. However, as in each failed attempt of analyzing simple algorithms, we were left with several smaller open problems, whose solution appeared still challenging but hopefully more achievable.

1.1.2.1. *The 3-Majority dynamics.* By trying to develop tools for analyzing LPA-like dynamics in order to solve the community detection problem, we ended up investigating majority dynamics and the results presented in Chapter 5. Very promising evidence in this direction was provided by [**DF11**], where it is proved that a dynamics not-too-far from those adopted in LPAs, the 3-Median dynamics, is extremely efficient in solving consensus problems even if there are a lot of initial labels in the system.

However, as outlined in Section 2.2, we surprisingly found that the convergence time in solving the consensus problem of the simplest majority dynamics, the 3-Majority process, is essentially linear in the number of initial different opinions in the system. We further proved that the situation does not change if instead of the 3-Majority we consider any protocol within a wide class of dynamics (h-input dynamics), and that the 3-Majority dynamics was already optimal w.r.t. all those dynamics which basically consist in exchanging opinions making use of at most 3 inputs (we may call such class LPA *with arity 3*).

These results were very bad news for the potential use of 3-Majority dynamics as a building block for more complex protocols and as an efficient dynamics per-se, and motivated the further investigation of faster dynamics for achieving plurality consensus. After exploring the vast space of possible candidates for quite a while, oscillating between dynamics which are no better than the 3-Majority dynamics and others whose analysis seems to be out of reach of current mathematical tools, we found ourselves in front of the Undecided-State dynamics, which is the subject of Chapter 6.

1.1.2.2. *The Undecided-State dynamics.* The Undecided-State dynamics was already famous in computer science as an elegant solution to more restricted majority consensus problems than the one we were considering in relation to LPAs. After some attempts at proving upper bounds on its convergence

[5]We remark that in applicative scenarios each color represents an opinion or more generally a class of a partition of the possible states of the agents.

time w.r.t. the standard hypotheses that are assumed in majority consensus problems, we discovered that under its deceptively simple structure the Undecided-State dynamics shows an evolution with an unexpected anatomy. Namely, its behaviour and convergence time are a function sensible to the *whole* initial configuration, instead of depending on few crucial parameters. We named this function the *monochromatic distance*. As discussed in Section 2.3, by inspecting the monochromatic distance we see that the Undecided-State dynamics has the advantage of having a convergence time which is at least as good as that of the 3-Majority dynamics (for a number of opinions in the system which can be as large as $\sqrt{n}/\log n$), and exponentially faster for a wide range of configurations. Thus, it is a simple but way more effective dynamics in many applicative contexts.

However, despite the sensible progress in analyzing dynamics that could serve as the core of a simple community detection protocol, midway through the author's PhD, the day in which we could be able to come up with a provably effective dynamics for community detection seemed quite far. At some point, Luca Trevisan suggested to look at the Averaging dynamics, which have the advantage of being linear and thus analyzable using the tools of spectral graph theory. Stepping away from LPA-based protocols turned out to be the right move: by developing a new analysis of the famous Averaging dynamics, we were finally able to prove that such a simple dynamics can efficiently solve the community detection problem.

1.1.2.3. *The Averaging dynamics.* By leveraging on the fact that, in a precise sense, the Averaging dynamics is *implicitly simulating* the calculation of the second eigenvector via a matrix power method, our analysis allows the definition of a simple labeling scheme that, on top of the Averaging dynamics, performs a global clustering on a wide class of graphs whose cluster structure is sufficiently reflected on the second eigenvector of their adjacency matrix. The latter class notably includes the famous stochastic block model, which has attracted a lot of attention as an interesting mathematical object to investigate the computational hardness of community detection. We show that the efficiency of the Averaging dynamics is comparable to that of the best, centralized and sophisticated techniques. As discussed in Section 2.1, this result provides one of the few examples of a *dynamics* [**AAE08, AAB+11, Dot14, MNT14**] that solves a computational problem that is non-trivial in a centralized setting.

Despite its simplicity, the Averaging dynamics still has the disadvantages of assuming that agents can interpret their state as a real number and perform arithmetical operations. Furthermore, the dynamics operates in the $\mathcal{LOCAL}$ model [**Pel00**]. Therefore, the quest for a simpler LPA-based dynamics for community detection remains open, as discussed in Chapter 10.

While chapters 4, 5 and 6 are dedicated to the analysis of specific dynamics, chapters 8 and 9 are devoted to the application of the 3-Majority

dynamics and a variant of it. In the remainder of this section we outline the motivation that led us to study the subjects of chapters 7, 8 and 9.

1.1.2.4. *Parallel random walks in the $\mathcal{PUSH}$ model.* The analysis in Chapter 6 strongly relies on the complete topology of the underlying interaction graph, i.e. on the fact that the $\mathcal{PULL}$ model is unstructured: all pairs of agents have the same interaction capability. The direct analysis of the Undecided-State dynamics on sparser topologies is a challenging open problem. However, in the $\mathcal{GOSSIP}$ model in which nodes are constrained to interact with only one neighbor but the communication can be bidirectional and nodes can choose with whom they interact, it is possible to leverage on the power of this dynamics for the $\mathcal{PULL}$ model even when the interaction topology has good expansion and regularity (see Section 2.3.1 for formal details). In fact, on good regular expander graphs[6] it is possible to efficiently simulate the $\mathcal{PULL}$ model in the $\mathcal{GOSSIP}$ model via a simple random-walk-based strategy.

The crucial issue that affects random walks in the $\mathcal{GOSSIP}$ model is that the model constrains each node to initiate at most one interaction per round. Consequently, if several random walks happen to be on the same node, they are not able to move away from it onto different neighbors at the same time (see Figure 23). The latter issue generates some *congestion*. In Section 6.3 of Chapter 6, we show that in the given setting the congestion is negligible at the cost of a small factor in the running time, provided that the random walks are required to run for few rounds. However, whether the congestion of random walks in the $\mathcal{GOSSIP}$ model remains small even when the topology doesn't exhibit good expansion, or when the random walks need to reach a considerable length, is not known and, as discussed in Section 2.4, there is strong evidence that these problems require major advances with respect to the available techniques.

Chapter 7 presents a modest attempt to make progress in understanding the congestion of random walks in the uniform $\mathcal{PUSH}$ model[7] by considering the behaviour of the process on the complete graph in the long run. Observe that the operations of reception and dispatch of tokens by which nodes implement random walks on the graph, are simple operations which

[6]Recall that an expander graph $G = (V, E)$ is a graph whose edge expansion is lower bounded by a constant, i.e.

$$h(G) = \min_{0<|S|\leqslant \frac{n}{2}} \frac{|E(S, V-S)|}{|S|},$$

where $E(A, B) := \{(u, v) \in E : u \in A, v \in B\}$ for $A, B \subseteq V$.

[7]As pointed out in Section 2.4, when we are only interested in the behavior of random walks in the $\mathcal{GOSSIP}$ model without the need to perform other operations (e.g. rewinding the random walks as in the simulation of $\mathcal{PULL}$ model in Section 6.3), we do not need to assume that nodes can control with whom they interact or that they can request information from the contacted node, unless we want to consider more complicated ways of implementing the random walks, but the latter attempt would lead us too far from a dynamics.

satisfy the requisites of dynamics (provided the tokens are not too many, to keep a low memory requirement for nodes). Similarly, the nodes can implement random walks in the uniform $\mathcal{PUSH}$ model by equipping themselves with a FIFO queue, which still produces a dynamics (observe that the larger memory requirement is compensated by a very limited communication capability). In fact, proving that with high probability (*w.h.p*[8] for short), the nodes' FIFO queues do not exceed a small size is the practical goal of analyzing the random walk process. In the aforementioned setting, we show that the congestion does not depart significantly from that of classical parallel random walks in the $\mathcal{LOCAL}$ model (see Section 2.4 for the formal statements). As a byproduct, in Section 7.3 we get an efficient dynamics for the problem of *parallel resource assignment* in the uniform $\mathcal{PUSH}$ model.

1.1.2.5. *Noisy bit dissemination and plurality consensus.* In addition to the purely theoretical interest and potential applications in technological contexts (e.g. sensor and ad-hoc networks), this work is also partially motivated by biological questions (chapters 8 and 9). Indeed, in the biological world, bit dissemination and majority consensus are a common phenomenon in a wide range of systems. Examples of such processes include a single ant that has found food and recruits others [**REF13, HW90**], few cells that trigger large population responses [**FJT+10**], a school of fish that reaches consensus around a group of leaders [**SKJ+08**], or a small number of observant individuals that alert their herd [**Rob96**]. Such information propagation is achieved despite what appears to be highly unpredictable, uncoordinated, noisy and limited communication settings. How biological systems manage to operate effectively despite such communication limitations is a fundamental question whose understanding is still very preliminary.

The previous research direction was a tempting ground for the author when, at the end of his first year of PhD, he had the pleasure of being Pierre Fraigniaud's guest at the computer science lab LIAFA[9]. At that time the author had concluded the work on the Undecided-State dynamics [**BCN+15a**], whose hardness originates from dealing with the setting in which the number of opinions in the system can be a function of the system size (see Section 2.3). We decided to work on the generalization of a work by Amos Korman et al. to the setting with multiple possible opinions. In Korman et al.'s work they investigate "natural" protocols for solving the bit dissemination and majority consensus problems in a noisy version of the uniform $\mathcal{PUSH}$ model [**FHK14**] (see Section 2.5), where each message can be corrupted (in fact, *changed*), before being received.

We thus began the research that led us to the results presented in Chapter 8. Our generalization required us to solve both conceptual and technical issues.

---

[8]We say that a sequence of events $\mathcal{E}_n$, $n = 1, 2, \ldots$ holds *with high probability* if $\Pr(\mathcal{E}_n) = 1 - \mathcal{O}(1/n^\gamma)$ for some positive constant $\gamma > 0$.

[9]LIAFA was later renamed IRIF.

On the conceptual side, while in the binary-message case the noise merely consists in the fact that with some probability one of the two values can be "flipped", in the multivalued case it is not clear what is the *right* way of modeling the fact that messages can be misunderstood. Here, the "right" modeling is the formalization that allows to separate in the most natural way the settings in which the problem can be solved from those in which it is not solvable. In Chapter 8, we provide a natural formalization of the noise and identify some crucial properties which allows a precise characterization of the solvability of the problems at hand.

On the technical side, the problem shares the following usual difficulty of generalizing a finite-volume process from dimension one to more than one. In the binary case, informally speaking, what is not 1 has to be 0: the fact that there are only two possible values provides the possibility to "take the complement" of quantities regarding one value, to get those regarding the other one. This possibility, which is often a key ingredient of the analysis, vanishes when we introduce further degrees of freedom in the process by allowing more than two possible values. Furthermore, the generalization to the multivalued case worsens the stochastic dependency that is already affecting the binary one, preventing a direct application of standard concentration-of-probability inequalities. As a byproduct of the analysis presented in Chapter 8, we provide a general framework to eliminate such dependencies.

While the presented generalization in the end is still far from being a dynamics per-se, the rules which the whole protocol is based-on are not: the core of the algorithm in fact relies on a generalization of the 3-Majority dynamics.

1.1.2.6. *Self-stabilizing majority bit dissemination.* As Pierre Fraigniaud's guest at LIAFA in Paris, the author was delighted to meet Amos Korman, who was working on his ERC proposal on "Distributed Biological Algorithms". The common interest in applying distributed computing ideas to understanding biological systems was soon evident. We briefly recall Amos Korman's observation on the biological significance that the consensus problem has in nature, which he expressed in one of the first conversations with the author.

While in a technological setting reaching consensus is often seen as the pre-condition for achieving some other goal, in a biological setting maintaining consensus is an evolutionary convenient strategy to cope with the limitations of single individuals in acquiring information from the environment (e.g. in answering questions such as "Is there a predator around?"), and to maximize the probability of survival in general (e.g. isolated individuals are easier preys). Thus, the tendency for a biological system to reach consensus is more of an instinct instilled by evolution than a behavior consciously adopted to achieve another agenda.

As an example, let us imagine a group of birds on a wire. At some point some bird starts to fly. The other birds have the legitimate doubt that the

moving one is leaving her spot on the wire because she has caught sight of a predator. Therefore, other birds start flying as well. Perhaps, shortly after leaving her point on the wire the first bird lands again on it, since her original intention was only to move to a better place. The other alarmed birds then realize that it was a false alarm, and they also start landing again on the wire. On the other hand, the first bird may also continue her escape from an imminent threat, which causes more and more other birds to leave the wire as they see other fellows doing it, and even the most distracted one rapidly realizes that it might be wiser to take off.

From the previous anecdotal example, we can abstract the following distributed computing problem. We have a system of agents in the $\mathcal{PULL}$ model (see Section 2.2), and one of them, the *source*, has some important piece of information that the system could use, which we call *input bit*[10]. However, there is no assumption on the initial states of the agents: some of them, for example, may hold a wrong assumption on the value of the input bit. Therefore, we would like to devise a strategy, as simple as possible, such that the system can rapidly reach consensus on the true value of the input bit, starting from an arbitrary initial configuration of the agents' states. In particular, we would like the system to converge fast to a configuration in which all agents are aware of the value of the input bit, and to be fast in updating the agents' knowledge of the input bit whenever the source changes her mind. In the terminology of distributed computing, we would like a solution which is a *self-stabilizing* protocol (see Definition 9).

Given the previous abstract formulation, we are essentially asked to solve the self-stabilizing consensus problem in the setting in which there is one agent (the source) which *does not change her mind* (she *knows* the true input bit). Thus, given the above anecdotal motivation for the problem, Amos Korman informally referred to the problem as the "stubborn bird" problem.

After a year of work with Amos and his student Lucas Boczkowski, we were able to leverage on the power of simple dynamics and prove the results outlined in Section 2.6. There, we illustrate the sound connection of the self-stabilizing bit dissemination problem with the problem of synchronizing clocks, in a self-stabilizing manner, in the uniform $\mathcal{PULL}$ model. We thus end up devising a solution for the self-stabilizing clock synchronization problem. The protocol we present in Chapter 9 uses, as a subroutine, any dynamics for majority consensus such as the 3-Median dynamics and, in the uniform $\mathcal{PULL}$ model using messages of 3 bits only, the presented solution allows the agents to synchronize a clock modulo T in time essentially logarithmic in T and the size of the system.

As showed in Chapter 9, this allows to remove the assumption of an initial common time notion from an entire class of protocols (defined in

[10]For simplicity's sake, we are assuming that the source's information is a binary value, i.e. a bit.

Section 2.6), and provides a general solution for the self-stabilizing *majority* bit dissemination problem, which is a generalization of the aforementioned bit dissemination problem which includes the majority consensus problem as a special case.

We have concluded an outline of the story behind the results proved in this work. In the following chapter we present them formally, with a detailed discussion on their meaning and significance.

CHAPTER 2

Overview of Results

In this chapter, we thoroughly discuss the obtained results that are then proved in the successive chapters, following the same order of topics. We thus begin with the Averaging dynamics which, as we show, is able to solve a computational problem (the community detection problem) which is non-trivial even in a centralized setting, thus making a strong case for the computational power of dynamics.

2.1. Distributed Community Detection via Averaging

Consider the following distributed algorithm on an undirected graph, which we call Averaging protocol[1]. At the outset, every node picks an initial value, independently and uniformly at random in $\{-1, 1\}$; then, in each synchronous round, every node updates its value to the average of those held by its neighbors. A node also tags herself "blue" if the last update increased its value, "red" otherwise. (See also the pseudocode in Algorithm 1.)

Averaging protocol

Rademacher initialization: At round $t = 0$ every node $v \in V$ independently samples its value from $\{-1, +1\}$ uniformly at random;

Updating rule: At each subsequent round $t \geqslant 1$, every node $v \in V$

(1) (Averaging dynamics) Updates its value $\mathbf{x}^{(t)}(v)$ to the average of the values of its neighbors at the end of the previous round

(2) (Coloring) If $\mathbf{x}^{(t)}(v) \geqslant \mathbf{x}^{(t-1)}(v)$ then v sets $\texttt{color}^{(t)}(v) = \texttt{blue}$, otherwise v sets $\texttt{color}^{(t)}(v) = \texttt{red}$.

Algorithm 1. Pseudocode of the Averaging protocol.

In Chapter 4, we prove that under various graph models exhibiting sparse balanced cuts (definitions 3, 4, 5, 6), including the *stochastic block model* (Definition 6 at page 32, see also Section 3.1.2) [**HLL83**], the process resulting from the above simple local rule converges, in logarithmic time, to a coloring that reflects the underlying cut, either exactly or approximately

[1]Note that the names "Averaging protocol" and "Averaging dynamics" denotes different protocols: the latter is the update function applied in step (1) of the updating rule of the former (see Algorithm 1).

depending on the graph model. The case of an exact identification of the two communities is called a *strong* reconstruction, while the case of an approximate identification of the cut is called a *weak* reconstruction, as stated in the following definition.

DEFINITION 2 (Strong and Weak Reconstruction). Given a graph $G = (V_1 \cup V_2, E)$ with $V_1 \cap V_2 = \emptyset$, a *weak (block) reconstruction* is a two-coloring of the nodes that separates V_1 and V_2 up to a small fraction of the nodes. Formally, we define an *ε-weak reconstruction* as a map

$$f : V_1 \cup V_2 \to \{\texttt{red}, \texttt{blue}\}$$

such that there are two subsets $W_1 \subseteq V_1$ and $W_2 \subseteq V_2$ with[2]

$$|W_1 \cup W_2| \geqslant (1-\varepsilon)|V_1 \cup V_2| \quad \text{and} \quad f(W_1) \cap f(W_2) = \emptyset.$$

When $\varepsilon = 0$ we say that f is a *strong reconstruction*.

Finally, we further show that our approach simply and naturally extends to more communities, providing a quantitative analysis for a regularized version of the stochastic block model with multiple communities. A roadmap of the main results is given in Figure 2.

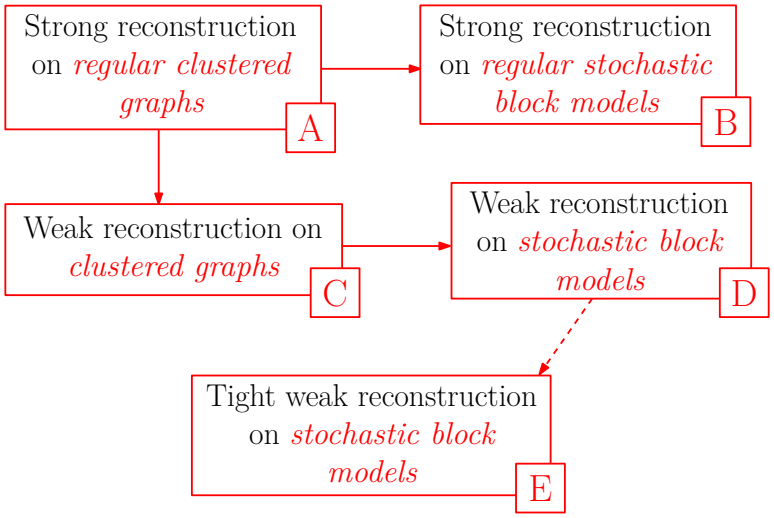

FIGURE 2. Summary of the results proved in Chapter 4: *A)* $\to$ Theorem 1, *B)* $\to$ Theorem 3, *C)* $\to$ Corollary 1, *D)* $\to$ Corollary 2, *E)* $\to$ Theorem 2.

More precisely, consider a graph $G = (V, E)$. We show that, if a partition (V_1, V_2) of G exists, such that $\mathbf{1}_{V_1} - \mathbf{1}_{V_2}$ is[3] (or is close to) a right-eigenvector

[2]We adopt the common convention that $f(S) := \{f(x) : x \in S\}$ for any function f with domain D and any subset $S \subseteq D$.

[3]As explained further, $\mathbf{1}_{V_i}$, is the vector with $|V|$ components, such that the j-th component is 1 if $j \in V_i$, it is 0 otherwise.

of the second largest eigenvalue of the transition matrix of G, and the gap between the second and the third largest eigenvalues is sufficiently large, our algorithm identifies the partition (V_1, V_2), or a close approximation thereof, in a logarithmic number of rounds. Though the Averaging dynamics does not explicitly perform any eigenvector computation, it exploits the spectral structure of the underlying graph: in some sense, the dynamics is an *implicit* distributed simulation of the power method.

The presented analysis involves two main novelties, relating to how nodes assign themselves to clusters, and to the spectral bounds that we prove for certain classes of graphs. A conceptual contribution is to make each node, at each round t, assign herself to a cluster ("find its place") by considering the difference between its value at time t and its value at time $t-1$. Such a criterion removes the component of the value lying in the first eigenspace without explicitly computing it. This idea has two advantages: it allows a particularly simple algorithm, and it gives a running time that depends on the third eigenvalue of the transition matrix of the graph. In graphs that have the structure of two expander graphs[4] joined by a sparse cut, the running time of the dynamics depends only on the expansion of the components and it is faster than the mixing time of the overall graph (see Figure 12). As discussed in Section 3.1, the Averaging dynamics is the first distributed reconstruction algorithm converging faster than the mixing time.

The Averaging dynamics works on any graph where

- the right-eigenspace of the second eigenvalue of the transition matrix is correlated to the cut between the two clusters and
- the gap between the second and third eigenvalues is sufficiently large.

While these conditions have been investigated for the spectrum of the *adjacency* matrix of the graph, the analysis of the Averaging protocol requires these conditions to hold for the *transition* matrix. A technical novelty of the analysis in Chapter 4 is to show that such conditions are met by a large class of graphs, that includes graphs sampled from the *stochastic block model*. Proving spectral properties of the transition matrix of a random graph is more challenging than proving such properties for the adjacency matrix, because the entries of the transition matrix are not independent[5].

In the following sections we discuss in detail individual results on the specific models we consider.

2.1.1. Strong reconstruction for regular clustered graphs

In Section 4.4, we consider a $(2n, d, b)$-clustered regular graph G with adjacency matrix A, where the clustered regular graphs are the following broad family of instances whose regularity allows us to provide a particularly clean analysis.

[4]Recall the definition of expander graph in footnote 6 on page 22.

[5]See the proof of Lemma 11 for further details.

DEFINITION 3 (Clustered Regular Graph). A $(2n, d, b)$-clustered regular graph $G = ((V_1, V_2), E)$ is a connected graph over node set $V_1 \cup V_2$, with $|V_1| = |V_2| = n$ and such that:

- Every node has degree d,
- Every node in cluster V_1 has b neighbors in cluster V_2 and every node in V_2 has b neighbors in V_1.

If the two subgraphs induced by V_1 and V_2 are good expander graphs[6] and b is sufficiently small, the second and third eigenvalues of the graph's transition matrix $P = (1/d) \cdot A$ are separated by a large gap. In this case, we prove that the following happens *w.h.p.*[7]: If the Averaging dynamics is initialized by having every node choose a value uniformly and independently at random in $\{-1, 1\}$, within a logarithmic number of rounds the system enters a regime in which nodes' values are monotonically increasing or decreasing, depending on the community they belong to (see Figure 12). As a consequence, every node can apply a simple and completely local clustering rule in each round, which eventually results in a strong reconstruction. Formally, we thus prove the following, where

$$\lambda = \max\{|\lambda_3|, |\lambda_{2n}|\}$$

is the largest eigenvalue of P other than λ_1 and λ_2.

THEOREM 1 (Strong Reconstruction). *Let $G = ((V_1, V_2), E)$ be a connected $(2n, d, b)$-clustered regular graph with $1 - 2b/d > (1 + \delta)\lambda$ for an arbitrarily-small constant $\delta > 0$. Then the Averaging protocol produces a strong reconstruction within $\mathcal{O}(\log n)$ rounds, w.h.p.*

We then show that, under mild assumptions, a graph selected from the following *regular stochastic block model* [**BDG+16**] is a $(2n, d, b)$-clustered regular graph that satisfies the above spectral gap hypothesis, w.h.p.

DEFINITION 4 (Regular Stochastic Block Model). In the regular stochastic block model with two communities, a graph on $2n$ nodes is obtained as follows: Given two parameters $a(n)$ and $b(n)$ (*internal* and *external* degrees, respectively), partition nodes into two equal-sized subsets V_1 and V_2 and then sample a random $a(n)$-regular graph over each of V_1 and V_2 and a random $b(n)$-regular graph between V_1 and V_2.

REMARK 2. The regular stochastic block model can be instantiated in different ways depending on how one samples the random regular graphs (for example, via the uniform distribution over regular graphs, or by taking the disjoint union of random matchings) [**MNS14, BDG+16**].

We thus obtain a fast and extremely simple dynamics for strong reconstruction, over the full range of parameters of the regular stochastic block

[6] Recall the definition of expander graph in footnote 6 on page 22.

[7] Recall the meaning of w.h.p. as in footnote 8 on page 23.

model for which this is known to be possible using centralized algorithms [**MNS14, BDG+16**].

COROLLARY 1 (Reconstruction in Regular Stochastic Block Models). *Let G be a random graph sampled from the regular stochastic block model with*

$$a - b > 2(1+\eta)\sqrt{a+b}$$

for an arbitrarily small constant $\eta > 0$, then the Averaging protocol produces a strong reconstruction in $\mathcal{O}(\log n)$ rounds, w.h.p.

We further show that a natural extension of the Averaging protocol, in which nodes maintain an array of values and an array of colors, correctly identifies a hidden balanced k-partition in a regular clustered graph with a gap between eigenvalues λ_k and λ_{k+1}.

THEOREM 2 (More Communities). *Let $G = (V, E)$ be a k-clustered d-regular graph defined as above and assume that*

$$\lambda = \max\{|\lambda_{2n}|, \lambda_{k+1}\} < (1-\varepsilon) \cdot \frac{a-b}{d},$$

for a suitable constant $\varepsilon > 0$. Then, for $\ell = \Theta(\log n)$, the Averaging protocol with ℓ parallel runs produces a strong reconstruction within $\mathcal{O}(\log n)$ rounds, w.h.p.

We remark that graphs sampled from the regular stochastic block model with k communities satisfy the conditions of Theorem 2, w.h.p.

2.1.2. Weak reconstruction for non-regular clustered graphs

In Section 4.5, we extend the analysis of Section 4.4 on regular graph models to show that the Averaging dynamics also ensures weak reconstruction in clustered graphs having two clusters that satisfy an approximate regularity condition, according to the following definition, and that also exhibit a gap between second and third eigenvalues of the transition matrix P.

DEFINITION 5 (Clustered γ-Regular Graphs). A $(2n, d, b, \gamma)$-clustered graph $G = ((V_1, V_2), E)$ (with $\gamma < 1$), is a graph over node set $V_1 \cup V_2$, where $|V_1| = |V_2| = n$ such that:

- Every node has degree $d \pm \gamma d$,
- Every node in V_1 has $b \pm \gamma d$ neighbors in V_2 and every node in V_2 has $b \pm \gamma d$ neighbors in V_1.

Given a $(2n, d, b, \gamma)$-clustered graph, in Chapter 4 we prove the following result.

THEOREM 3 (Weak Reconstruction). *Let G be a connected $(2n, d, b, \gamma)$-clustered graph with $\gamma \leqslant c(\nu - \lambda_3)$ for a suitable constant $c > 0$. If $\lambda < \nu$ and $\lambda_2 \geqslant (1+\delta)\lambda$ for an arbitrarily-small positive constant δ, then the Averaging*

protocol produces an $\mathcal{O}(\gamma^2/(\nu-\lambda_3)^2)$*-weak reconstruction within* $\mathcal{O}(\log n)$ *rounds, w.h.p.*[8]

As an application, we then prove that these conditions are met by the *stochastic block model*, which offers a popular framework for the probabilistic modelling of graphs that exhibit good clustering or community properties (see Section 3.1.2 for a discussion of the significance of the model). We here consider the following simple version with two communities of equal size.

DEFINITION 6 (Stochastic Block Model). The stochastic block model $\mathcal{G}_{2n,p,q}$, a.k.a. planted bisection model, consists of $2n$ nodes and an edge probability distribution defined as follows: The node set is partitioned into two subsets V_1 and V_2, each of size n; edges linking nodes belonging to the same partition appear in E independently at random with probability $p = p(n)$, while edges connecting nodes from different partitions appear with probability $q = q(n) < p$ (see also Figure 3).

Calling $a = pn$ and $b = qn$, we prove that graphs sampled from $\mathcal{G}_{2n,p,q}$ satisfy w.h.p. the above approximate regularity and spectral gap conditions of Theorem 3, whenever $a - b > 25\sqrt{(a+b)\cdot \log n}$ (Lemma 7), thus proving the following result.

COROLLARY 2 (Reconstruction in Stochastic Block Models). *Let* $G \sim \mathcal{G}_{2n,p,q}$*. If* $a - b > 25\sqrt{d \log n}$ *and* $b = \Omega(\log n/n^2)$ *then the Averaging protocol produces an* $\mathcal{O}(d \log n/(a-b)^2)$*-weak reconstruction in* $\mathcal{O}(\log n)$ *rounds w.h.p.*

We remark that the latter result for the stochastic block model follows from an analysis that applies to general *non-random* clustered graphs and hence does not exploit crucial properties of random graphs. A further technical contribution described in Chapter 4 is a refined, ad-hoc analysis of the Averaging dynamics for the $\mathcal{G}_{2n,p,q}$ model, showing that this protocol achieves weak-reconstruction in logarithmic time whenever $a - b > \Omega_\varepsilon(\sqrt{(a+b)})$.

THEOREM 4 (Tight Reconstruction in Stochastic Block Models). *Let* $G \sim \mathcal{G}_{2n,p,q}$*. If*

$$(a-b)^2 > c_{\text{opt}}(a+b) > 5\log n,$$

and[9] $a + b < n^{\frac{1}{3}-c_{\text{tight}}}$ *for some positive constants* c_{opt} *and* c_{tight}*, then the Averaging protocol produces an* $\mathcal{O}(d/(a-b)^2)$*-weak reconstruction within* $\mathcal{O}(\log n)$ *rounds w.h.p.*

This refined analysis requires a deeper understanding of the eigenvectors of the *transition matrix* of G. Coja-Oghlan [**CO10**] defined certain graph properties that guarantee that a near-optimal bisection can be found based on eigenvector computations of the *adjacency matrix*. Similarly, we show

[8] Consistently, Theorem 1 is a special case of this one when $\gamma = 0$.

[9] It should be possible to weaken the condition $d < n^{\frac{1}{3}-c_{\text{tight}}}$ via some stronger concentration argument; see the proof of Lemma 16 at the end of the chapter for details.

simple sufficient conditions under which a right eigenvector of the second largest eigenvalue of the transition matrix of a graph approximately identifies the hidden partition. We give a tight analysis of the spectrum of the transition matrix of graphs sampled from the stochastic block model in Section 4.7. Notice that the analysis of the transition matrix is somewhat harder than that of the adjacency matrix, since the entries are not independent of each other; we are not aware of comparable results in the existing literature, which we review in Section 3.2.

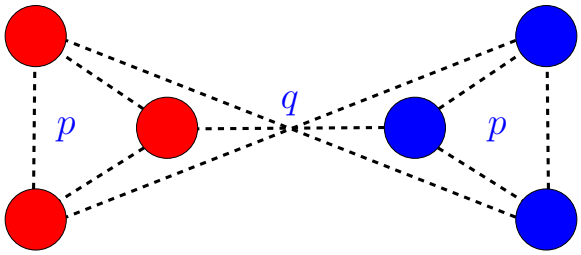

FIGURE 3. A representation of the stochastic block model (Definition 6): edges linking nodes belonging to the same community are included in the graph independently at random with probability $p = p(n)$, while edges connecting nodes in different communities are included with probability $q = q(n) < p$.

2.1.3. Beyond the Averaging dynamics: a wrap up

The results presented in Section 2.1 show rigorous evidence of the possibilities offered by completely decentralized, extremely simple and natural dynamics to address computational problems that are complex even in a centralized setting, such as community detection in clustered graphs, whose complexity appears far beyond most of the tasks to which this kind of dynamics have been traditionally applied in the area of distributed computing.

However, we remark that the Averaging dynamics is a linear dynamics, which requires the nodes to be able to hold rational values and to perform basic arithmetic operations on them. Furthermore, the Averaging dynamics operates in the $\mathcal{LOCAL}$ model, in which each node at each round can send and receive a message from each neighbor. As discussed in Section 1.1.2, inspired by the empirical success of label propagation algorithms, in the next two sections we are going to investigate simpler dynamics, which are non-linear and operate in random sparse communication models[10]. As a

[10]In the rest of this work we are going to consider more restrictive models such as the $\mathcal{GOSSIP}$ model, the stochastic restriction of the $\mathcal{GOSSIP}$ model known as uniform $\mathcal{GOSSIP}$ model, and the unidirectional restrictions of the uniform $\mathcal{GOSSIP}$ model known as uniform $\mathcal{PULL}$ and $\mathcal{PUSH}$ models. In the next section, we always consider the $\mathcal{PULL}$ model. In all these models, as is remarked in the following chapters, the interactions among nodes are very sparse: typically each node interacts with very few neighbors.

consequence, they are way more efficient in terms of communication cost and way more robust.

Our understanding of the behavior of non-linear dynamics is still at its infancy and is not sufficient to allow a rigorous analysis of their sophisticated uses, e.g. for community detection. In order to get to the point in which there is reasonable hope to carry on such rigorous analyses, we first have to understand their behavior in solving more basic problems. This is the purpose of sections 2.2 and 2.3, in which we study two dynamics for some consensus problems, being fundamental issues that naturally arise as sub-problems of more complex tasks such as community detection.

2.2. The 3-Majority Dynamics: Plurality and Stabilizing Consensus

In this section and the next one we consider the *stabilizing consensus* and the *plurality consensus* problems in the context of a communication network in which each of n anonymous nodes supports an initial opinion chosen from a finite set $[k]$, which we can think of as colors. We first consider the plurality consensus problem, in which the initial hypothesis of an initial bias toward the plurality opinion allows to circumvent some core difficulties of the general consensus problem (See Section 5.4.1).

2.2.1. The 3-Majority dynamics for plurality consensus

In the plurality consensus problem it is assumed that the initial (opinion) configuration has a sufficiently large *bias* s towards a fixed opinion $m \in [k]$ - that is, the number c_m of nodes supporting the plurality opinion (in short, the *initial plurality size*) exceeds the number c_j of nodes supporting any other opinion j by an additive value s. The goal is to design an efficient fully-distributed protocol that let the network converge to the *plurality consensus*, i.e., to the monochromatic configuration in which all nodes support the plurality opinion.

Reaching plurality consensus in a distributed system is a fundamental problem arising in several areas such as distributed computing [**DGM$^+$11, Pel02**], communication networks [**PVV09**], and social networks [**CDIG$^+$13, MS10, MNT14**]. Following some works analyzing dynamics for this problem [**AD15, DGM$^+$11**] (which are reviewed in Section 3.3), we study the 3-Majority dynamics, which is a discrete-time, synchronous process in which, at every round, each of the n anonymous nodes samples independently and uniformly at random three nodes[11], including herself and with repetitions, and adopts the plurality opinion among those three (breaking ties uniformly at random). We consider one of the simplest models, the uniform $\mathcal{PULL}$ model, in which the network is a clique.

[11] We remark that looking at only two random nodes and breaking ties uniformly at random would yield a process equivalent to the *polling process* [**HP01**] (see Lemma 66), which is known to converge to a *minority* opinion with constant probability even for $k = 2$ and large initial bias (i.e. $s = \Theta(n)$) [**HP01**].

In [**DGM$^+$11**], a tight analysis of a 3-input dynamics for the *median* problem on the clique was presented: the goal there is to converge to a stable configuration where all nodes support a value which is a good approximation of the *median* of the initial configuration. It turns out that, in the binary case (i.e $k = 2$), the median problem is equivalent to plurality consensus and the 3-input dynamics for the median is equivalent to the 3-Majority dynamics: As a result, they obtain, for any bias $s \geqslant c\sqrt{n \log n}$ for some constant $c > 0$, an optimal bound $\Theta(\log n)$ on the convergence time of the 3-Majority dynamics for the binary case of the problem considered here. However, for any $k \geqslant 3$, it is easy to see that the two problems above differ significantly (in particular, the median may be very different from the plurality) and thus, the two dynamics are different from each other as well. Moreover, the analysis in [**DGM$^+$11**] - strongly based on the properties of the median function - cannot be adapted to bound the convergence time of the 3-Majority dynamics.

Previously to the results presented in this section and proved in Chapter 5, the role of the parameter $k = k(n)$ (the number of initial opinions), in the convergence time of this dynamics was unknown and, more generally, the existence of efficient dynamics reaching plurality consensus for $k \geqslant 3$ was left as an important open issue in [**AAE08, DGM$^+$11, BD13**]. In Chapter 5, we present an analysis of the 3-Majority dynamics in the general case (i.e. for any $k \in [n]$). A consequence of such analysis which exemplifies the results of Chapter 5 is the following.

Corollary 3 (Upper Bound with Bias). *Let* $\mathbf{c}$ *be any initial* k*-color configuration with*

$$s(\mathbf{c}) \geqslant 72\sqrt{2\min\left\{2k,\ \sqrt[3]{\frac{n}{\log n}}\right\} n \log n}.$$

Then, the 3*-Majority dynamics converges to the plurality opinion in* $\mathcal{O}(\min\{2k, \sqrt[3]{n/\log n}\}\log n)$ *time w.h.p.*

The proof technique in Section 5.2 is accurate enough to get another interesting form of the above upper bound that does not depend on k. In fact, Corollary 3 is a particular case of the following general theorem.

Theorem 5 (General Upper Bound for 3-Majority). *Let* λ *be any value such that* $\lambda < \sqrt[3]{n}$ *and let* $\mathbf{c}$ *be any initial* k*-cd, with* $c_1 \geqslant n/\lambda$ *and*

$$s(\mathbf{c}) \geqslant 72\sqrt{2\lambda\, n \log n}.$$

Then the 3*-Majority dynamics converges to the plurality opinion in* $\mathcal{O}(\lambda \log n)$ *time w.h.p.*

In particular, Theorem 5 implies that the convergence time is polylogarithmic when the size of the plurality opinion is of order $n/\text{polylog}n$, as follows.

COROLLARY 4 (Polylogarithmic Upper Bound for 3-Majority). *Let* $\mathbf{c}$ *be any initial* k*-cd with* $c_1 \geqslant n/\log^{\ell} n$ *and*

$$s(\mathbf{c}) \geqslant 72\sqrt{2n\log^{\ell+1} n}.$$

Then, the 3*-Majority dynamics converges to the plurality opinion in* $\mathcal{O}(\log^{\ell+1} n)$ *time w.h.p.*

We then show that the upper bound of Theorem 5 is tight for a wide range of the input parameters. When $k \leqslant (n/\log n)^{1/4}$, we prove the following lower bound $\Omega(k \log n)$ on the convergence time of the 3-Majority dynamics.

THEOREM 6 (Lower Bound for 3-Majority). *Let*

$$\tau = \inf\{t \in \mathbb{N} : \mathbf{C}^{(t)} \text{ is monochromatic}\}$$

be the random variable indicating the first round such that the system is in a monochromatic configuration. If the initial number of opinions is $k \leqslant (n/\log n)^{1/4}$ *and the initial configuration is* $\mathbf{c} = (c_1, \ldots, c_k)$ *with*

$$\max\{c_j : j = 1, \ldots, k\} \leqslant \frac{n}{k} + \left(\frac{n}{k}\right)^{1-\varepsilon}$$

for some $\varepsilon > 0$*, then* $\tau = \Omega(k \log n)$ *w.h.p.*

Observe that the range of k in Theorem 5 largely includes the initial bias required by our upper bound when $k \leqslant (n/\log n)^{1/4}$. So, the *linear-in-k* dependence of the convergence time cannot be removed for a wide range of the parameter k.

The analysis presented in Chapter 5 provides a clear picture of the 3-Majority dynamic process. Informally speaking, the larger the initial value of c_m is (w.r.t. n), the smaller the required initial bias s and the faster the convergence time are. On the other hand, the lower-bound argument shows, as a by-product, that the initial plurality size c_m needs $\Omega(k)$ rounds just to increase from $n/k + o(n/k)$ to $2\,n/k$.

We then prove a general negative result: Under the distributed model we consider, within the class of dynamics using no additional state other than the initial opinions, no dynamics with at most 3 inputs (other than the 3-Majority dynamics) converges w.h.p. to plurality consensus starting from any initial configuration with bias $s = o(n)$. The latter result is formally stated in Theorem 21. The statement requires few definitions (definitions 10, 11, 12 and 14), and is deferred to Section 5.3.2.

In other words, within the class above, not only there is no 3-input dynamics that achieves convergence to plurality consensus in $o(k \log n)$ rounds, but the 3-Majority dynamics is the only one that eventually achieves this goal at all, no matter how long the process takes. Rather interestingly, by comparing the $\mathcal{O}(\log n)$ bound for the 3-Median dynamics [**DGM+11**] to our negative results for the plurality on the same distributed model, we get an exponential time-gap between the task of computing the median and the

one of computing plurality (this happens for instance when $k = n^a$, for any constant $0 < a < 1/4$).

A natural question suggested by the previous results is whether (slightly) larger random samples of nodes' neighborhoods might lead to significant improvements in convergence time to plurality consensus. We provide a negative answer to this question. To this purpose, we consider the h-Plurality dynamics, i.e., the natural generalization of the 3-Majority dynamics in which every node, in each round, updates her opinion according to the plurality of the opinions supported by h randomly sampled neighbors. We prove the following lower bound.

THEOREM 7 (Lower Bound for h-Majority). *Let $\mathbf{C}^{(t)}$ be the random variable indicating the configuration at round t according to the h-Plurality dynamics and let*

$$\tau = \inf\{t \in \mathbb{N} : \mathbf{C}^{(t)} \text{ is monochromatic}\}.$$

If the initial configuration $\mathbf{c} = (c_1, \ldots, c_k)$ is such that

$$\max\{c_j : j = 1, \ldots, k\} \leqslant \frac{3n}{2k},$$

then $\tau = \Omega(k/h^2)$ w.h.p.

We emphasize that scalable and efficient protocols must yield low communication complexity and small node congestion in every round. These properties are guaranteed by the h-Plurality dynamics only when h is small, say $h = \mathcal{O}(\text{polylog}(n))$: In this case, our lower bound implies that the resulting speed-up is only polylogarithmic with respect to the 3-Majority dynamics.

One motivation for adopting dynamics in reaching (*simple*) consensus[12] (such as the 3-Median dynamics in [**DGM+11**]) lies in their provably-good *self-stabilizing* properties against *dynamic adversary corruptions*: It turns out that the 3-Majority dynamics has good self-stabilizing properties for the *plurality consensus* problem. More formally, a *T-bounded adversary* knows the state of every node at the end of each round and, based on this knowledge, she can corrupt the opinion of up to T nodes in an arbitrary way, just before the next round begins. In this case, the goal is to achieve an almost-stable phase where all but at most $\mathcal{O}(T)$ nodes agree on the plurality value. This "almost-stability" phase must have $\text{poly}(n)$ length, with high probability. Our analysis shows that the 3-Majority dynamics guarantees the self-stabilization property for plurality consensus, as given in the following.

COROLLARY 5 (Upper Bound with Adversary). *Let λ be any value such that $\lambda < \sqrt[3]{n}$ and let $\mathbf{c}$ be any initial configuration, with $c_1 \geqslant n/\lambda$ and*

$$s(\mathbf{c}) \geqslant 24\sqrt{2\lambda\, n \log n}.$$

[12] In the (simple) consensus problem the goal is to reach any stable monochromatic configuration (any opinion is accepted) starting from any initial configuration.

The 3-Majority dynamics achieves $\mathcal{O}(s(\mathbf{c})/\lambda)$-plurality consensus against any F-bounded adversary with $F = o(s(\mathbf{c})/\lambda)$, and the convergence time is $\mathcal{O}(\lambda \log n)$ w.h.p.

We have concluded our overview of the results proved in sections 5.1, 5.2 and 5.3 concerning the performance of the 3-Majority dynamics in solving the plurality consensus problem. In the next section, we basically drop the fundamental assumption made so far, that is the presence of an initial *bias* between the plurality opinion and all other ones. Rather than converging to a specific value, as we explain shortly our next goal is to converge to any opinion in a *stable* way.

2.2.2. The 3-Majority dynamics for stabilizing consensus

Let us call Σ the finite set of possible initial opinions. We call an opinion *valid* if it is held by at least one node at the beginning.

In this section we are interested in the following scenario: After every node performs a step of 3-Majority dynamics by pulling the opinion from three random nodes and setting her new opinion to the majority one (breaking ties arbitrarily), an adaptive *dynamic adversary* can arbitrarily change the opinions of a subset of the nodes, possibly choosing different subsets over different rounds (see Figure 4). We consider *F-dynamic adversaries* that, at every round, can change the opinions of up to F nodes, possibly introducing non-valid opinions.

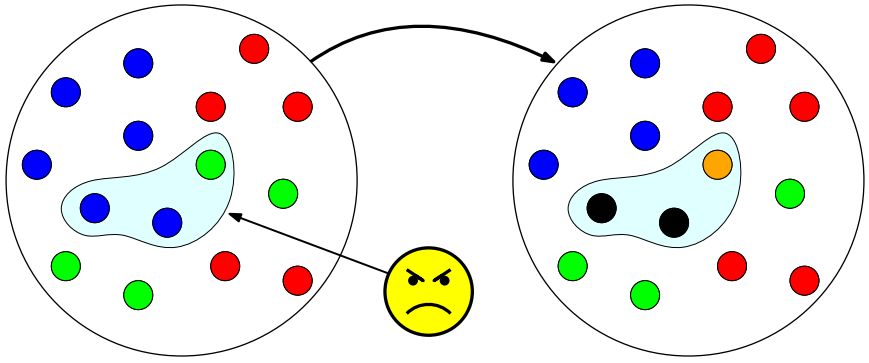

FIGURE 4. At the end of each round, an F-dynamic adversary can change the opinions of F nodes, possibly choosing different subsets of nodes over different rounds.

Let the system start from any configuration having k valid opinions with $k \leqslant n^{\alpha}$ for a suitable constant $\alpha < 1$ and consider any F-dynamic adversary with $F = \mathcal{O}(\sqrt{n}/(k^{5/2} \log n))$. We prove that the process converges to a configuration in which all but $O(\sqrt{n})$ nodes hold the same valid opinion within $O((k^2\sqrt{\log n} + k \log n)(k + \log n))$ rounds, w.h.p. (see Theorem 8 below). This shows that the 3-Majority dynamics provides an efficient solution

to the *stabilizing-consensus* problem in the uniform $\mathcal{PULL}$ model. Previously to our result, this was known only for the binary case, i.e. $|\Sigma| = 2$, while for any $|\Sigma| \geqslant 3$, it has been an important open question for several years [**AAE08, DGM+11**]. Furthermore, still for any $|\Sigma| \geqslant 3$, $o(n)$-time convergence of the 3-Majority dynamics was open even in the absence of an adversary whenever the initial bias toward some plurality opinion is not large.

In this section we describe in more detail the consensus problem and various network scenarios in which it is of interest, and the results in this setting proved in Chapter 5, while we defer a comparison with previous related results to Section 3.5.

2.2.2.1. *Consensus (or Byzantine agreement).* The *consensus* problem in a distributed network is defined as follows: A collection of agents, each holding a piece of information (an element of a set Σ), interact with the goal of agreeing on one of the elements of Σ initially held by at least one agent, possibly in the presence of an adversary that is trying to disrupt the protocol. The consensus problem in the presence of an adversary (known as Byzantine agreement) is a fundamental primitive in the design of distributed algorithms [**PSL80, Rab83**]. The goal is to design a distributed, local protocol that brings the system into a configuration that meets the following conditions:

(1) *Agreement*: All non-corrupted nodes support the same opinion v;
(2) *Validity*: The opinion v must be a *valid* one, i.e., an opinion which was initially declared by at least one (non-corrupted) node;
(3) *Termination*: Every non-corrupted node can correctly decide to stop running the protocol at some round.

There is considerable interest in the design of consensus algorithms in models that severely restrict both communication and computation [**AAE08, BCN+15a, DGM+11**], both for efficiency considerations and because such models capture aspects of the way consensus is reached in social networks, biological systems, and other domains of interest in network science [**AAD+06, AFJ06, BSDDS10, CCN12, Dot14, FHK14, FPM+02**].

As in the previous section, we consider the uniform $\mathcal{PULL}$ model. In this paragraph, we briefly review the model and the underlying assumptions. In compliance with the requirements of dynamics, we consider the above problem in the restrictive setting of an anonymous network in which nodes possess no unique IDs, nor do they have any static binding of their local link ports (i.e., nodes cannot keep track of *who sent what*). From the point of view of computation, the most prohibitive setting is to assume that each node only has $\mathcal{O}(\log |\Sigma|)$ bits of memory available, i.e., it barely suffices to store the number of opinions. We further assume that this bound extends to link bandwidth available in each round. Finally, communication capabilities are severely constrained and non-deterministic: Every node can communicate with at most a (small) constant number of random neighbors in each round. These constraints are well-captured by the uniform $\mathcal{PULL}$

communication model [**DGH$^+$87, KSSV00, KDG03**]: At every round, every node can exchange a (short) message (say, $\Theta(\log(|\Sigma|))$ bits) with each of at most h random neighbors, where h is a (small) absolute constant[13]. A sequential variant of the uniform $\mathcal{PULL}$ model is the *(random) population-protocols* model [**AAE08, AAE06, AAD$^+$06**] in which, in each round, a single interaction between a pair of randomly selected nodes occurs.

The classic notion of consensus is too strong and unrealistic in the aforementioned distributed settings, that instead rely on *weaker* forms of consensus, deeply investigated in [**AAE08, AFJ06, Asp12, DGM$^+$11**]. In this chapter, we consider a variant of the *stabilizing-consensus* problem [**AFJ06**] considered in [**AAE08**]: There, a solution is required to converge to a stable *regime* in which the above three properties are guaranteed in a relaxed, still useful form[14]. More precisely:

Definition 7 (Stabilizing Almost-Consensus). Starting from any initial configuration with k valid opinions, a *stabilizing almost-consensus* protocol must ensure the following properties:

- *Almost agreement.* In a finite number of rounds, the system must reach a *regime* of configurations where all but a *negligible* "bad" subset (i.e. having size $\mathcal{O}(n^\gamma)$ for constant $\gamma < 1$) of the nodes support the same opinion.
- *Almost validity.* The system is required to converge w.h.p. to an almost-agreement regime where all but a negligible bad set of nodes keep the same *valid* opinion.
- *Non termination.* In dynamic distributed systems, nodes represent simple and anonymous computing units which are not necessarily able to detect any global property.
- *Stability.* The convergence toward such a weaker form of agreement is only guaranteed to hold with high probability[15] and only over a *long period* (i.e. for any arbitrarily-large polynomial number of rounds).

We remark that, prior to the results presented in this work, no stabilizing almost-consensus protocol was known for $|\Sigma| > 2$ even in the complete graph.

A major result in Chapter 5 is about the convergence properties of the 3-Majority dynamics in the uniform $\mathcal{PULL}$ model in the presence of the adaptive F-dynamic adversary defined above.

Theorem 8 (Upper Bound with Dynamic-Adversary). *Let $k \leqslant n^\alpha$ and $F \leqslant \beta\sqrt{n}/(k^{\frac{5}{2}} \log n)$ for some constants $\beta, \alpha > 0$. The 3-Majority dynamics*

[13] In fact, $h = 1$ in the standard uniform $\mathcal{PULL}$ model. It is easy to verify that all our results still hold in this more restricted model at the cost of a constant slow-down in convergence time and local memory size.

[14] These relaxed convergence properties are described in detail in Section 7 of [**AAE08**].

[15] Recall the meaning of w.h.p. as in footnote 8 on page 23.

is a stabilizing almost-consensus protocol in the presence of any F-dynamic adversary and its convergence time is $\mathcal{O}((k^2\sqrt{\log n} + k\log n)(k+\log n))$, w.h.p.

As a simple consequence of the analysis provided in Chapter 5, we also get the following bound on the convergence time in the presence of any *F-static adversary* with a larger bound on F, where by *F-static adversary* we mean an adversary that looks at the initial configuration, then changes the opinion of up to F arbitrary nodes and, after that, no further adversary's actions are allowed.

COROLLARY 6 (Upper Bound with Static-Adversary). *Starting from any initial configuration with $k \leqslant n^{\alpha}$ active opinions, where $\alpha > 0$ is a suitable constant, the 3-Majority dynamics reaches almost-consensus within $\mathcal{O}((k^2\sqrt{\log n} + k\log n)\cdot(k+\log n))$ rounds, in the presence of any F-static adversary with $F \leqslant n/k - \sqrt{kn\log n}$, w.h.p.*

We remark that Theorem 6 provides an $\Omega(k\log n)$ bound on the convergence time of the 3-Majority dynamics, which holds even when the system starts from biased configurations.

Not assuming a large initial bias of the plurality opinion considerably complicates the analysis. Indeed, the major open challenge is the analysis from (almost) uniform configurations, where the system needs to break the initial symmetry in the absence of significant drifts towards any of the initial opinions. So far, the symmetry breaking in the 3-Majority dynamics has never been analyzed even in the non-adversarial case. Moreover, the phase before symmetry breaking is the one in which the adversary has more chances to cause undesired behaviours: Long delays and/or convergence towards non-valid opinions. In Section 5.4, after providing some preliminaries, we shall discuss the above technical challenges.

Finally, one may wonder whether it is possible to provide guarantees about the opinion that eventually achieves majority. As for this point, the results of Chapter 5 (lemmas 29 and 30 in Section 5.4.4) imply that an opinion is not going to become majority unless it is a *near-plurality*, i.e. it is close to the size of the plurality opinion.

2.3. The Undecided-State dynamics: Plurality Consensus

In this section, we consider the Undecided-State dynamics[16] that has been introduced in [**AAE08**] and analyzed in [**AAE08, PVV09**] only in the binary case (i.e. $k = 2$). The analysis of the multivalued case (i.e. $k > 2$) has been proposed in [**AAE08, AD15, CER14, DGM$^+$11, MRSDZ11, JKV12**] as an open problem. The interest for this dynamics touches areas beyond the borders of computer science. It appears to play a major role

[16]The Protocol has been initially "designed" for the case $k = 2$ and, thus, in previous works it has been named the *Third-State* Dynamics.

in important biological processes modelled as so-called chemical reaction networks [**CCN12, Dot14**].

$u \backslash v$	undecided	opinion i	opinion j
undecided	undecided	i	j
i	i	i	undecided
j	j	undecided	j

TABLE 1. The update rule of the Undecided-State dynamics where $i, j \in [k]$ and $i \neq j$.

In this chapter we analyze the synchronous version of the dynamics in the (uniform) $\mathcal{PULL}$ model:

Undecided-State dynamics

AGENTS' POSSIBLE STATES: Each agent either supports an opinion $i \in [k]$ or she is in the *undecided* state, an extra state that agents can support. The undecided state does not count as an opinion, and agents supporting it are said to be undecided (or equivalently, to have no opinion).

1: u pulls the state of a randomly-selected neighbor v.
2: If u is supporting any opinion, and v's opinion differs from u's one, the agent enters the *undecided* state. Note that u does not update her state if her opinion coincides with v's one.
3: If u is undecided, she copies v's state.

Algorithm 2. One round of Undecided-State dynamics, executed by each agent u. (see also Table 1.)

We investigate the efficiency of Undecided-State dynamics w.r.t. the plurality consensus problem. As in Section 2.2, recall that in the plurality consensus problem each agent of a distributed system initially supports an opinion, i.e. a number $i \in [k] = \{1, 2, \ldots, k\}$ (with $2 \leqslant k \leqslant n$). In the initial opinion configuration $\mathbf{c} = (c_1, \ldots, c_k)$ (where c_i denotes the number of agents supporting opinion $i \in [k]$), there is an initial *plurality* c_1 of agents supporting the *plurality opinion* (w.l.o.g., we assume that opinion communities are ordered, so that $c_i \geqslant c_{i+1}$ for any $i \leqslant k-1$). Initially, every agent only knows her own opinion; the goal is to find a distributed algorithm that, w.h.p.[17], brings the system into the *target* configuration, i.e., the monochromatic configuration in which all agents support the initial plurality opinion. In the remainder, the subset of agents supporting opinion i is called the *i-opinion community*.

As discussed further in Section 3.4, the performance of Undecided-State dynamics on the complete graph has been evaluated w.r.t. the following

[17]Recall the meaning of w.h.p. as in footnote 8 on page 23.

parameters: the number n of nodes, the number k of opinions, and the initial *bias* towards the plurality opinion, with the latter characterized in terms of a parameter that only depends on the relative magnitude[18] of c_1 and c_2.

However, when $k > 2$, any such measure of the initial bias is not sensitive enough to accurately capture the convergence time of a plurality protocol: a *global* measure is needed, i.e., one that reflects the whole initial opinion configuration. To better appreciate this issue, consider the two configurations $\mathbf{c}$ and $\mathbf{c}'$ in Figure 5. Whether the absolute difference or the relative ratio is used to measure the initial bias, the opinion configuration $\mathbf{c}'$ appears to be not "worse" than $\mathbf{c}$. Still, computer simulations and intuitive arguments suggest that, under any "natural" plurality protocol, the almost-uniform opinion distribution $\mathbf{c}'$ can result in much larger convergence time than the highly-concentrated opinion configuration $\mathbf{c}$.

To the best of our knowledge, the analysis presented in Chapter 6 is the first one which investigates the impact of the whole initial opinion configuration on the speed of convergence of plurality protocols.

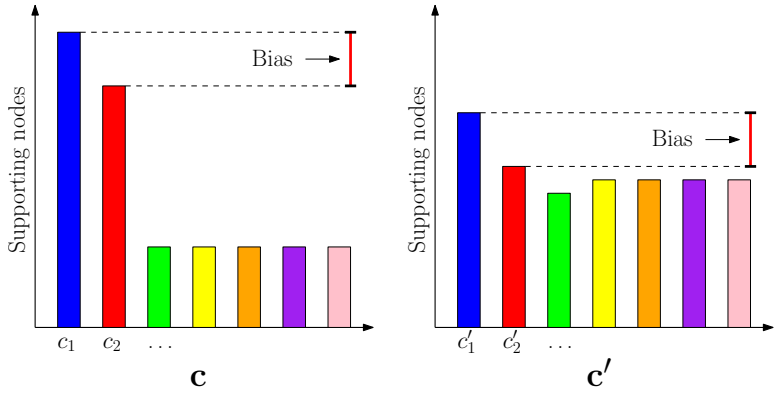

FIGURE 5. Two different opinion configurations having the same bias $s = s(c_1, c_2)$.

The core contribution of the analysis is represented by the introduction of a suitable distance $\mathsf{d}(\cdot,\cdot)$ (see Section 6.4.1 for a formal definition) on the set $\mathcal{S}$ of all opinion configurations. Such distance naturally induces a function $\mathsf{md}(\cdot)$, called the *monochromatic distance*, which equals the *distance* between any configuration $\mathbf{c}$ and the target configuration.

DEFINITION 8 (Monochromatic Distance). Given an opinion configuration $\mathbf{c}$, its monochromatic distance is defined as

$$\mathsf{md}(\mathbf{c}) \;=\; \sum_{i=1}^{k} \left(\frac{c_i}{c_1}\right)^2,$$

[18] Typically, this relative magnitude is defined in terms of the absolute difference or the ratio.

where c_1 is (one of) the plurality opinion(s).

We use md to characterize the bias of the initial configuration. In particular, note that $\mathsf{md}(\mathbf{c})$ measures the extent to which $\mathbf{c}$ is "uniform": Indeed, the higher the extent of the bias towards a small subset of the opinions (including the plurality one), the smaller the value of $\mathsf{md}(\mathbf{c})$. As an example, in Figure 5, $\mathsf{md}(\mathbf{c})$ can be substantially smaller than $\mathsf{md}(\mathbf{c}')$. At the extremes, when there are only $O(1)$ opinion communities of size $\Theta(c_1)$, we have $\mathsf{md}(\mathbf{c}) = \Theta(1)$ while, when $\Theta(k)$ opinion communities have size $\Theta(n/k)$, we have $\mathsf{md}(\mathbf{c}) = \Theta(k)$. A visual representation of md is provided in Figure 6.

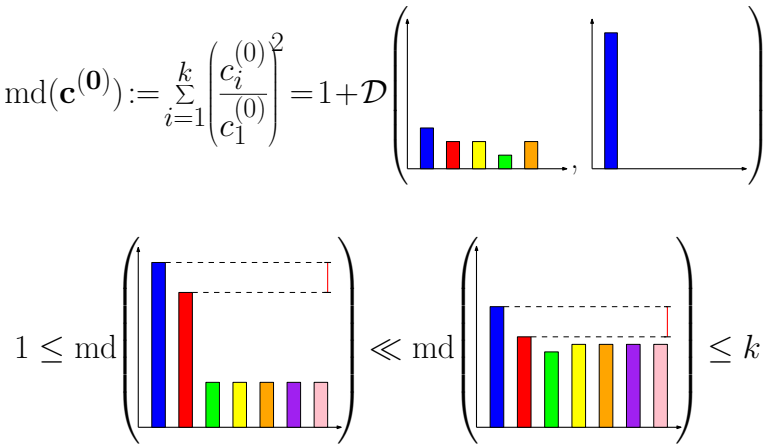

FIGURE 6. A visual representation of the monochromatic distance. At the extremes, when there are only $O(1)$ opinion communities of size $\Theta(c_1)$, we have $\mathsf{md} = \Theta(1)$ while, when $\Theta(k)$ opinion communities have size $\Theta(n/k)$, we have $\mathsf{md} = \Theta(k)$.

The simple strategy of the Undecided-State dynamics [**AAE08, PVV09**] is to "add" one extra state to somewhat account for the "previous" opinion supported by an agent (see Section 6.1 and Table 1 for a definition of this dynamics). In [**AD15, AAE08, BD13, BTV09, DV12, PVV09, JKV12**], the same dynamics has been analyzed under different distributed models and/or under very different initial assumptions (among others, under the assumption that k is an absolute constant). In these settings, important aspects of the complex dependence of the dynamics' evolution on the overall shape of the initial opinion configuration are missed.

We analyse the Undecided-State dynamics using a technique that strongly departs from past work and that allows us to address the plurality consensus problem in the general setting. Our analysis achieves almost-tight bounds on convergence time, as formally given by the following.

THEOREM 9 (Monochromatic Upper Bound). *Let $k = O\left((n/\log n)^{1/3}\right)$ and let $\mathbf{c}$ be any initial configuration such that $c_1 \geqslant (1+\alpha)\cdot c_2$ where α is an arbitrarily small positive constant. Then within time $O\left(\mathsf{md}(\mathbf{c})\cdot\log n\right)$ the system converges to the plurality opinion, w.h.p.*

This result is almost-tight in a strong sense, as expressed in the other following theorem.

THEOREM 10 (Monochromatic Lower Bound). *Let $k = O\left((n/\log n)^{1/6}\right)$. Starting from any opinion configuration $\mathbf{c}$ the convergence time of the Undecided-State dynamics is $\Omega(\mathsf{md}(\mathbf{c}))$, w.h.p.*

Let us compare Theorem 9 with the corresponding results in the previous section. According to Theorem 5 and Theorem 6, when the initial difference bias is $s = \Omega(\sqrt{kn\log n})$, the 3-Majority dynamics converges in $\Theta(\min\{k, n^{1/3}\}\log n)$ rounds using $\Theta(\log k)$ memory and message size. Convergence times of the 3-Majority dynamics become polylogarithmic only if $c_1 \geqslant n/\mathrm{polylog}(n)$, thus they are not polylogarithmic whenever $k = \omega(\mathrm{polylog}(n))$ and $c_1 = o(n/\mathrm{polylog}(n))$. This is the parameter range where analysis of the Undecided-State dynamics in Chapter 6 leads to an exponential speed up w.r.t. the convergence time of the 3-Majority dynamics. For example, consider an initial "oligarchic" scenario where $k = n^{1/4}$ and a subset $\mathcal{L} \subseteq [k]$ exists such that

- $|\mathcal{L}| = \mathrm{polylog}(n)$,
- for any $i \in \mathcal{L}$, $\bar{c}_i \sim n/\sqrt{k}$, and
- for any $i \in [k] \setminus \mathcal{L}$, $\bar{c}_i \sim n/k$.

Clearly, $1, 2 \in \mathcal{L}$ and the resulting monochromatic distance is $\mathsf{md}(\mathbf{c}) = \mathrm{polylog}(n)$. Assuming $c_1 \geqslant (1+\alpha)c_2$ for some $\alpha > 0$ the upper bound of Theorem 9 implies that, starting from any such configuration, the Undecided-State dynamics converges in polylogarithmic time, whereas the 3-Majority dynamics converges in $\Theta(k\log n)$ time (theorems 5 and 6).

2.3.1. Uniform $\mathcal{PULL}$ Simulation in the $\mathcal{GOSSIP}$ Model

The analysis of the Undecided-State dynamics provided in Chapter 6 is rather general and it can be extended to other interesting topologies. As a case supporting this claim, we show how to adapt the Undecided-State dynamics for the class of *d-regular expanders* [**HLW06**], for any degree $d \geqslant 1$.

In this variant of the Undecided-State dynamics, the task of selecting random neighbors is simulated by performing n independent random-walks of suitable length. Thanks to the well-known rapidly-mixing properties of d-regular expander graphs[19] [**HLW06, LPW09**], we can prove the following theorem.

THEOREM 11 (Monochromatic Bound on Expanders). *Let $G = (V, E)$ be a d-regular graph with constant expansion. For any initial configuration*

[19] Recall the definition of expander graph in footnote 6 on page 22.

c *such that the Undecided-State dynamics on the clique computes plurality consensus in* $O(\mathsf{md}(\mathbf{c}) \log n)$ *rounds w.h.p., the modified Undecided-State dynamics computes plurality consensus on* G *in* $O(\mathsf{md}(\mathbf{c}) \operatorname{polylog}(n))$ *rounds, w.h.p.*

The major technical hurdle here is proving that this variant of the protocol still requires $\operatorname{polylog}(n)$ local memory. To this aim, we prove that the *node congestion* is at most $\operatorname{polylog}(n)$. The analysis of the process that results from running parallel random walks over a graph has been the subject of extensive research in the past [**AAK$^+$08, FKP11, HPP$^+$12, Pel00, DSMP12**]. However, to the best of our knowledge, none has addressed the issues we consider here. In particular, the analysis of node congestion is far from trivial and of independent interest, since efficient protocols for several important tasks in the $\mathcal{GOSSIP}$ model (such as *node-sampling* [**DSMP12**], *network-discovery* problems [**HPP$^+$12**], and *averaging* problems [**BGPS06**]) rely on the use of parallel random walks. This leads us directly to the subject of the next section, which is the study of random walks in the uniform $\mathcal{PUSH}$ model.

In the next section we depart from the specific application of random walks in the $\mathcal{GOSSIP}$ model which is instrumental to Theorem 11, and we study the congestion that affects the dynamics which results by running parallel random walks in the uniform $\mathcal{PUSH}$ model, as an important primitive also to other problems discussed in the next section.

2.4. Random Walks in the $\mathcal{PUSH}$ Model

In this section we study the execution of n parallel random walks in the uniform $\mathcal{PUSH}$ model, in which at each round each node can send a message to a neighbor chosen uniformly at random. We focus on the case of a complete graph.

In the setting of a complete topology, it is convenient to express the process as the following *repeated balls-into-bins* process. Given any $n \geqslant 2$, we initially assign n balls to n bins in an arbitrary way. Then, at every round, from each non-empty bin one ball is chosen according to some strategy (random, FIFO, etc) and re-assigned to one of the n bins uniformly at random.

It is easy to see that the latter process is equivalent to the former one, and that the fact that from each node (bin) only one token (ball) can move (be extracted) generates some stochastic dependence among the positions of the tokens (balls) and the number of tokens on each node (balls in each bin), i.e. the *maximum load* of the process. The objective of Chapter 7 is indeed to investigate the impact of the stochastic dependence on the maximum load.

More formally, inspired by previous notions of (load) stability [**AKU05, BFG03**], we study the maximum number of balls inside one bin at round t and we are interested in the largest *maximum load* $M^{(t)}$ achieved by the

process over a period of *any polynomial* length. We say that a configuration is *legitimate* if its maximum load is $\mathcal{O}(\log n)$ and a process is *stable* if, starting from any legitimate configuration, it only takes on legitimate configurations over a period of poly(n) length, w.h.p. We remark that this notion of stability is a probabilistic relaxation of the notion of *closure* required by self-stabilization, which asks that starting from any legitimate configuration, the process only takes on legitimate configurations (see Figure 7).

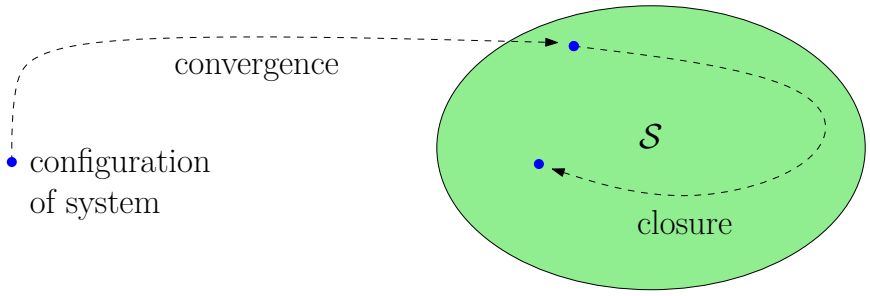

FIGURE 7. An illustration of the requirements of self-stabilization. Given the set $\mathcal{S}$:= {"legitimate configurations of the system" }, the protocol is required to guarantee two properties. The first one is **convergence**: From *any* initial configuration, the system has to reach $\mathcal{S}$. The second one is **closure**: If in $\mathcal{S}$, the system keep staying in $\mathcal{S}$. If we only require the two previous property to hold w.h.p., we get the definition of *probabilistic* self-stabilization (in which closure is called **stability**).

We formally define the probabilistic version of self-stabilization [**Dij74, Dol00**], as follows.

DEFINITION 9 ((Probabilistic) Self-Stabilizing Process). We say that a process is (stochastically) *self-stabilizing* if it is stable and if, moreover, starting from *any* configuration, it converges to a legitimate configuration, w.h.p. The *convergence time* of a self-stabilizing process is the maximum number of rounds required to reach a legitimate configuration starting from any configuration.

This natural notion of (probabilistic) self-stabilization has also been inspired by that in [**IJ90**] for other distributed processes.

Stability has consequences for other important aspects of this process. For instance, if the process is stable, we can get good upper bounds on the *progress* of a ball, namely the number of rounds the ball is selected from its current bin queue, along a sequence of $t \geqslant 1$ rounds (such implication is crucial in many applications, e.g. in Section 6.3). Furthermore, we can eventually bound the *parallel* cover time, i.e., the time required for every ball

to visit *all* bins. Self-stabilization has also important consequences when the system is prone to transient faults [**Dij74, Lam85, Dol00**].

The repeated balls-into-bins process was first studied in [**BCEG10**], where it is used as a crucial sub-procedure to optimize the message complexity of a gossip algorithm in the complete graph, and then in [**BCN$^+$15a, EK15**]. The analysis in [**BCEG10, EK15**] (only) holds for very-short (i.e. logarithmic) periods, while the analysis given in Section 6.3 considers periods of arbitrary length but it (only) allows to achieve a bound on the maximum load that rapidly increases with time: after t rounds, the maximum load is bounded by $\mathcal{O}\big(\sqrt{t}\big)$, w.h.p. By adopting the FIFO strategy at every bin queue, the latter result easily implies that the progress of any ball over periods of t rounds is $\Omega(\sqrt{t})$, w.h.p. On the other hand, an upper bound $\mathcal{O}\big(n^2 \log n\big)$ for the parallel cover time of the repeated balls-into-bins process easily follows from the fact that the cover time of one single random walk on the complete graph is $\Theta(n \log n)$, w.h.p.

Previous results are thus not helpful to establish whether this process is stable (or, even more, stochastically self-stabilizing) or not. Moreover, the previous analyses of the maximum load in [**BCN$^+$15a, BCEG10, EK15**] are far from tight, since they rely on some rough approximations of the studied process via other, much simpler Markov chains: for instance, in Chapter 6, we present the approach adopted in [**BCN$^+$15a**], in which they consider the process - which clearly dominates the original one - where, at every round, a new ball is inserted in every empty bin. That analysis thus does not exploit the global invariant (a fixed number n of balls) of the original process.

In Chapter 7, we provide the following, almost-tight analysis of the repeated balls-into-bins process that significantly departs from previous ones and show that the system is stochastically self-stabilizing.

THEOREM 12 (Repeated Balls into Bins Max Load). *Let c be an arbitrarily-large constant and let $\mathbf{q}$ be any legitimate configuration. Let the* repeated balls-into-bins process *start from $\mathbf{Q}^{(0)} = \mathbf{q}$. Then, over any period of length $\mathcal{O}(n^c)$, the process visits only legitimate configurations, w.h.p., i.e. $M^{(t)} = \mathcal{O}(\log n)$ for all $t = \mathcal{O}(n^c)$, w.h.p. Moreover, starting from any configuration, the system reaches a legitimate configuration within $\mathcal{O}(n)$ rounds, w.h.p.*

The previous result strongly improves over the best previous bounds [**BCN$^+$15a, BCEG10, EK15**] and it is almost tight, since the classical lower bound $\Omega(\log n / \log\log n)$ on the maximum load (see, e.g., [**MU05**]) clearly applies also in our repeated setting. Theorem 12 further implies that, under the FIFO queueing policy, any ball performs $\Omega(t/\log n)$ steps of its individual random walk over any sequence of $t = \text{poly}(n)$ rounds w.h.p., which implies that the parallel cover time is $\mathcal{O}\big(n \log^2 n\big)$, w.h.p. This is only a $\log n$ factor away from the lower bound following from the single-ball process.

2.4.1. An application to multiple resources assignment

We observe that the process of parallel random walks in the uniform $\mathcal{PUSH}$ model, models a natural randomized solution to the problem of *(parallel) resource (or task) assignment* in distributed systems (this problem is also known as *traversal*) [**San06, Lyn96**]. In the basic case, the goal is to assign one resource in mutual exclusion to *all* processors (i.e. nodes) of a distributed system. This is typically described as a *traversal* process performed by a *token* (representing the resource or task) over the network. The process terminates when the token has visited all nodes of the system. Randomized protocols for this problem [**Coo11**] are efficient approaches when, for instance, the network is prone to faults/changes and/or when there is no global labeling of the nodes. A simple randomized protocol is the one based on *random walks* [**Coo11, IJ90, IKOY02**]: starting from any node, the token performs a random walk over the network until all nodes are visited, w.h.p. The first round in which all nodes have been visited by the token is called the *cover time* of the random walk [**Coo11, LPW09**]. The expected cover time for general graphs is $\mathcal{O}(|V| \cdot |E|)$ (see, for example, [**MU05**]).

In distributed systems, we often are in the presence of *several* resources or tasks that must be processed by every node *in parallel*. This naturally leads to consider the parallel version of the basic problem in which n different tokens (resources) are initially distributed over the set of nodes and every token must visit all nodes of the network. Similarly to the basic case, an efficient randomized solution is the one based on (parallel) random walks. In order to visit the nodes, every token performs a random walk under the natural constraint that every node can process and release at most one token per round. Again, the maximum load is a critical complexity measure: for instance, it can determine the required buffer size at every node, bounds on the token progress and, thus, on the parallel cover time. For this case, our results imply that, every token visits all nodes of the system with at most a logarithmic delay w.r.t. the case of a single token: so, we can derive an upper bound $\mathcal{O}(n \log^2 n)$ for the parallel cover time, starting from *any* initial configuration. We can also consider the adversarial model in which, in some *faulty* rounds, an adversary can re-assign the tokens to the nodes in an arbitrary way. The self-stabilization and the linear convergence time shown in Theorem 12 imply that the $\mathcal{O}\big(n \log^2 n\big)$ bound on the cover time still holds, provided that faulty rounds occur with a frequency no higher than cn, for a sufficiently large constant c.

In the next sections we continue our exposition of applications of dynamics by investigating two basic problems in distributed computing, the bit dissemination (better known as *rumor spreading*) and the plurality consensus problems (the second of which has already been the main character of sections 2.2 and 2.3), in two challenging fundamental settings. In Section 2.5, we consider the problems in the uniform $\mathcal{PUSH}$ model when communication is affected by noise, i.e. when there is large chance that messages

sent are "misunderstood". In Section 2.6 we consider the problem of bit dissemination in the $\mathcal{PULL}$ model in the self-stabilizing context. We show that the self-stabilizing bit dissemination problem is deeply connected to that of clock-synchronization, and we thus investigate also the latter. The rationale behind the order of the two section is given by the increase in sophistication of the two solutions: in Section 2.5 (corresponding to Chapter 8), we provide an algorithm that, although not as simple as a dynamics, is still arguably *natural*; in Section 2.6 (corresponding to Chapter 9), the solution uses dynamics as a black box but, although the resulting protocol is simple from a technological point of view, it cannot be argued to be biologically relevant.

2.5. Bit Dissemination and Consensus Despite Noise

To guarantee reliable communication over a network in the presence of noise is the main goal of Network Information Theory [**EGK11**]. Thanks to the achievements of this theory, the impact of noise can often be drastically reduced to almost zero by employing *error-correcting codes*, which are practical methods whenever dealing with artificial entities. However, the situation is radically different for scenarios in which the computational entities are biological. Indeed, from a biological perspective, a computational process can be considered "simple" only if it consists of very basic primitive operations, and is extremely lightweight. As a consequence, it is unlikely that biological entities are employing techniques like error-correcting codes to reduce the impact of noise in communications between them. Yet, biological signals are subject to noise, when generated, transmitted, and received. This rises the intriguing question of how entities in biological ensembles can cooperate in presence of noisy communications, but in absence of mechanisms such as error-correcting codes.

An important step toward understanding communications in biological ensembles has been achieved in [**FHK14**], which showed how it is possible to cope with noisy communications in absence of coding mechanisms for solving complex tasks such as *bit dissemination* and *majority consensus*. Such a result provides highly valuable hints on how complex tasks can be achieved in frameworks such as the immune system [**Car04**], bacteria populations [**WB05**], or super-organisms of social insects [**HW09**], despite the presence of noisy communications.

In the case of bit dissemination we assume that a source-node initially handles a bit, set to some binary value, called the *correct opinion*. This opinion has to be transmitted to all nodes, in a noisy environment, modeled as a complete network with unreliable links. More precisely, messages are transmitted in the network according to the classical *uniform $\mathcal{PUSH}$ model* [**DGH$^+$87, KSSV00, Pit87**] where, at each round, every node can send one binary opinion to a neighbor chosen uniformly and independently at random but, before reaching the receiver, that opinion is flipped with probability at most $\frac{1}{2} - \varepsilon$ with $\varepsilon > 0$. We refer to this variant of uniform $\mathcal{PUSH}$ model

as the *noisy*[20] (uniform) $\mathcal{PUSH}$ model. In the case of majority consensus, it is assumed that some nodes are supporting opinion 0, some nodes are supporting opinion 1, and some other nodes are supporting no opinion. The objective is that all nodes eventually support the initially most frequent opinion (0 or 1). More precisely, let A be the set of nodes with opinion, and let $b \in \{0,1\}$ be the majority opinion in A. The *majority bias* of A is defined as $\frac{1}{2}(|A_b| - |A_{\bar{b}}|)/|A|$ where A_i is the set of nodes with opinion $i \in \{0,1\}$.

In [**FHK14**], it is proved that, even in above very noisy setting, the bit dissemination and the noisy majority consensus problems can be solved quite efficiently. Specifically, an algorithm is provided that solves the noisy bit dissemination problem in $O(\frac{1}{\varepsilon^2}\log n)$ communication rounds, with high probability[21] in n-node networks, using $O(\log\log n + \log(1/\varepsilon))$ bits of memory per node. Actually, as a special case of the previous algorithm, one gets an algorithm with the same aforementioned performances which solves the noisy majority consensus problem for $|A| = \Omega(\frac{1}{\varepsilon^2}\log n)$ with majority-bias $\Omega(\sqrt{\log n/|A|})$. Note that the provided majority consensus algorithm requires that the nodes are initially aware of the size of A. We remark that both algorithms exchange solely opinions between nodes, and are optimal, since basic information-theoretic arguments show that both bit dissemination and majority consensus require $\Omega(\frac{1}{\varepsilon^2}\log n)$ rounds in n-node networks, w.h.p.

Our objective here is to extend the work of [**FHK14**] to the natural case of an arbitrary number of opinions, to go beyond a proof of concept. The problem that results from this extension is an instance of the *plurality consensus* problem in the presence of noise, i.e., the problem of making the system converging to the initially most frequent opinion (i.e., the *plurality* opinion). Indeed, the plurality consensus problem naturally arises in several biological settings, typically for choosing between different directions for a flock of birds [**BSDDS10**], different speeds for a school of fish [**SKJ$^+$08**], or different nesting sites for ants [**FPM$^+$02**]. The computation of the most frequent value has also been observed in biological cells [**CCN12**].

The ultimate goal of our investigation is to make progress toward the solution of the above problems via simple dynamics. At present, the protocol of [**FHK14**] and that present here, although already very simple, are far from the time-homogeneous property of dynamics, since they rely on the ability of nodes to coordinate in adopting different rules at different times. However we remark that, *within* the single phases of these protocols, the mechanisms

[20] Observe that the smaller is ε, the more *uniformly random* received messages appear, and the problem becomes therefore harder. We remark that, even for very large values of ε the problem does not reduce to *adversarial* scenarios such as those considered in Section 2.2, or more general *byzantine* settings where even simple consensus cannot be achieved if the fraction of byzantine nodes exceeds $\frac{1}{3}$. For example, if $\varepsilon = \frac{1}{7}$, at each round a fraction greater than $\frac{1}{3}$ of the messages is corrupted, therefore a naive interpretation of corrupted messages as messages sent by adversarial agents is of no use.

[21] Recall the meaning of w.h.p. as in footnote 8 on page 23.

adopted by nodes essentially reduce to the h-Majority dynamics and other elementary rules.

We generalize the results in [**FHK14**] to the setting in which an arbitrary large number k of opinions is present in the system. In the context of bit dissemination, the correct opinion is a value $i \in \{1, \ldots, k\}$, for any constant $k \geqslant 2$. Initially, one node supports this opinion i, and the other nodes have no opinions. The nodes must exchange opinions so that, eventually, all nodes support the correct opinion i. We also recall that, as discussed in sections 2.2 and 2.3, in the context of (relative) majority consensus, also known as *plurality consensus*, each node u initially supports one opinion $i_u \in \{1, \ldots, k\}$, or has no opinion. The objective is that all nodes eventually adopt the *plurality opinion* (i.e., the opinion initially held by more nodes than any other, but not necessarily by an overall majority of nodes).

As in [**FHK14**], we restrict ourselves to "natural" algorithms [**Cha09**], which informally[22] means that the algorithm essentially consists in exchanging opinions in a straightforward manner (i.e., they do not use the opinions to encode, e.g., part of their internal state). For both problems, the difficulty comes from the fact that every opinion can be modified during its traversal of any link, and switched at random to any other opinion.

Generalizing noisy bit dissemination and noisy majority consensus to more than just two opinions requires to address a series of issues, some conceptual, others technical.

Conceptually, one needs first to redefine the notion of noise. In the case of binary opinions, the noise can just flip an opinion to its complement. In the case of multiple opinions, an opinion i subject to a modification is switched to another opinion i', but there are many ways of picking i'. For instance, i' can be picked uniformly at random (u.a.r.) among all opinions. Or, i' could be picked as one of the "close opinions", say, either $i+1$ or $i-1$ modulo k. Or, i' could be "reset" to, say, $i = 1$. In fact, there are very many alternatives, and not all enable bit dissemination and plurality consensus to be solved. One of our contributions is to characterize *noise matrices* $P = (p_{i,j})$, where $p_{i,j}$ is the probability that opinion i is switched to opinion j, for which these two problems are efficiently solvable. Similar issues arise for, e.g., redefining the majority bias into a *plurality bias*.

The technical difficulties are manifold. A key ingredient of the analysis in [**FHK14**] is a fine estimate of how nodes can mitigate the impact of noise by observing the opinions of *many* other nodes, and then considering the mode of such sample. Their proof relies on the fact that for the binary opinion case, given a sample of size γ, the number of 1s and 0s in the sample sum up to γ. Even for the ternary opinion case, the additional *degree of freedom* in the sample radically changes the nature of the problem, and the impact of noise is statistically far more difficult to handle.

[22] We are not aware of any serious attempt at a rigorous definition of what a natural algorithm is.

Also, to address the multivalued case, we have to cope with the fact that, in the uniform $\mathcal{PUSH}$ model, the messages received by nodes at every round are correlated. To see why, consider an instance of the system in which a certain opinion b is held by one node only, and there is no noise at all. In one round, only one other node can receive b. It follows that if a certain node u has received b, no other nodes have received it. Thus, the messages each node receives are not independent (see Figure 8). In Chapter 8, we show how to obtain concentration of probability in this dependent setting by leveraging Poisson approximation techniques. Our approach has the following advantage: instead of showing that the Chernoff bound can be directly applied to the specific involved random variables, we show that the execution of the given protocol, on the uniform $\mathcal{PUSH}$ model, can be tightly approximated with the execution of the same protocol over a suitable communication model, that is not affected by the stochastic correlation that affects the uniform $\mathcal{PUSH}$ model.

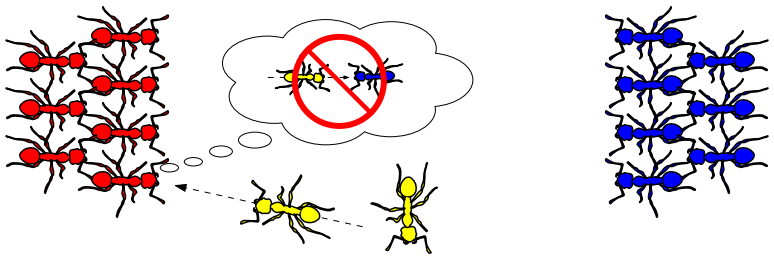

FIGURE 8. An example of the stochastic dependence which affects messages in the uniform $\mathcal{PUSH}$ model. The red ant which is contacted by the yellow one can infer that the probability that a blue ant is contacted by another yellow ant decreases.

In short, we prove that there are algorithms solving the noisy bit dissemination problem and the noisy plurality consensus problem for multiple opinions, with the same performances and probabilistic guarantees as the algorithms for binary opinions in [**FHK14**]. Below, we state the main theorems proved in Chapter 8, concerning the solution of the two problems. The statement require some notions which it would be too technical to rigorously provide here. In this introductory section, we only informally anticipate their meaning:

- A noise matrix is a matrix whose row i and column j give the probability of a message i to be changed to j by the noise, before being received;
- A δ-majority-biased configuration is a configuration of the system in which the most frequent opinion has a support of nodes which is larger than that of any other opinion by a fraction δ (Definition 19);

- A noise matrix is majority preserving (m.p., for short), with parameters ε and δ, if the probability that a message, sent by a randomly chosen node in a δ-majority-biased configuration, has the value of the majority opinion is larger than that of having any other opinion by at least $\varepsilon \cdot \delta$ (Definition 20).

Given the above notions, we prove the following results.

THEOREM 13 (Noisy Bit Dissemination). *Assume that the noise matrix P is (ε, δ)-m.p. with $\varepsilon = \Omega(n^{-\frac{1}{4}+\eta})$ for an arbitrarily small constant $\eta > 0$ and $\delta = \Omega(\sqrt{\log n / n})$. There exists a protocol, using $O(\log\log n + \log\frac{1}{\varepsilon})$ bits of memory at each node, which solves the noisy bit dissemination problem with k opinions in $O(\frac{\log n}{\varepsilon^2})$ communication rounds, w.h.p.*

THEOREM 14 (Noisy Plurality Consensus). *Let S with $|S| = \Omega(\frac{1}{\varepsilon^2}\log n)$ be an initial set of nodes with opinions in $[k]$, the rest of the nodes having no opinions. Assume that the noise matrix P is (ε, δ)-m.p. for some $\varepsilon > 0$, and that S is $\Omega(\sqrt{\log n / |S|})$-majority-biased. There exists a protocol, using $O(\log\log n + \log\frac{1}{\varepsilon})$ bits of memory at each node, which solves the noisy plurality consensus problem with k opinions in $O(\frac{\log n}{\varepsilon^2})$ communication rounds, w.h.p.*

In the last, next introductory section, we move from studying simple protocols for dealing with noise in the uniform $\mathcal{PUSH}$ model, to studying *transient faults* (or, in a biological perspective, *the effect of a dynamic environment*) in the uniform $\mathcal{PULL}$ model.

2.6. Self-Stabilizing Bit Dissemination

As in Section 2.5, the real-world scenario we consider in this section are distributed systems composed of limited agents that interact in a stochastic fashion to jointly perform tasks which are common in the natural world as well as in engineered systems, such as a wide range of insect populations [**HM85**], chemical reaction networks [**CCDS14**], and mobile sensor networks [**AAD$^+$06**]. Such systems have been studied in various disciplines, including biology, physics, computer science and chemistry, while employing different mathematical and experimental tools. For example, using computer simulations to model animal group interactions, Couzin et al. demonstrated how groups can reach majority-consensus decisions, even though informed individuals do not know whether they are in a majority or minority [**CKFL05**]. From an algorithmic perspective, such complex systems share a number of computational challenges. Indeed, they all perform collectively in dynamically changing environments despite being composed of limited individuals that communicate through seemingly unpredictable, unreliable, and restricted interactions.

In Section 2.5 (which introduces Chapter 8), we have focused on the unpredictability, unreliability and poorness of interactions of biological systems as abstracted by the noisy uniform $\mathcal{PUSH}$ model. The latter investigation takes part to the significant effort in understanding the computational limitations that are inherent to such systems, by abstracting some of their characteristics as distributed computing models, and analyzing them algorithmically [**AAD+06, AG15, DS15, FHK14, AFJ06, BCN+15a**]. As these models attempt to capture biological scenarios, they necessarily consider agents which are restricted in their memory and communication capacities, that interact independently and uniformly at random (u.a.r.). By now, the understanding of the computational power of such models is rather advanced. However, it is important to note that much of this progress has been made assuming *non-faulty scenarios* - a rather strong assumption when it comes to natural or sensor-based systems. For example, to synchronize actions between processors, many known distributed protocols rely on the assumption that processors know when the protocol is initiated. However, in systems composed of limited individuals that do not share a common time notion, and must react to a dynamically changing environment, it is often unclear how to achieve such conditions. To have a better understanding of such systems, it is desirable to identify the weakest computational models that still allow for both efficient as well as robust computations in a fault-tolerant sense.

In Chapter 9, we go back to the basic uniform $\mathcal{PULL}$ model of communication considered in chapters 5 and 6, in which in each round, each agent can extract (pull) information from few other agents, chosen u.a.r. In the computer science discipline, this model, as well as its companion $\mathcal{PUSH}$ model which we have considered in Section 2.5, gained their popularity due to their simplicity and inherent robustness to different kinds of faults [**DGH+88, KSSV00, DGM+11, DF11**]. Here, focusing more on the context of natural systems, we view the $\mathcal{PULL}$ model as an abstraction for communication in well-mixed scenarios, where agents can occasionally "observe" arbitrary other agents. This may relate to the notion of *passive communication* commonly used by biologists to refer to communication that is based on observing the behavior of other individuals [**Wil92**], in contrast to *active communication* in which agents "deliberately" signal other agents and whose corresponding model is the uniform $\mathcal{PUSH}$ model considered in Chapter 8.

We aim at identifying the power and limitations of the uniform $\mathcal{PULL}$ model with respect to achieving basic information dissemination tasks under conditions of increased uncertainty for the agents, regarding the state of the system they are in. As many natural systems appear to be more restricted by their communication abilities than by their memory capacities [**AAB+11, EW13**], our main focus is on understanding what can be computed while revealing as few bits per interaction as possible in a self-stabilizing way.

We note that stochastic communication patterns such as $\mathcal{PULL}$ or $\mathcal{PUSH}$ are inherently sensitive to congestion issues. Indeed, in such models it is unclear how to simulate a protocol that uses large messages while using only small size messages. For example, the straightforward strategy of breaking a large message into small pieces and sequentially sending them one after another does not work, since one typically cannot make sure that the small messages reach the same destination. Hence, reducing the message size may have a profound impact on the running time, and perhaps even on the solvability of the problem at hand.

Similarly to the previous section, here we consider the problem of disseminating information from one or several sources to the rest of the population, which is one of the most fundamental building blocks in distributed computing [**DGH+88, CHHKM12, DF11, KSSV00, CLP11**], and an important primitive in natural systems [**REF13, SKJ+08, Rob96**]. However, there are profound differences between the bit dissemination problem considered in Section 2.5 and the variant considered here: we consider the problem in the context of self-stabilization, and the generalization considered here includes the single-source bit dissemination and the majority consensus problems as special cases.

More formally, we focus on the *majority bit dissemination* problem defined as follows [**BKN17**]. We consider a population of n agents. The population may contain multiple *source agents* which are specified by a designated bit in the memory of every agent indicating whether the agent is a source or not. Each source agent holds a binary *input bit*, however, two sources may not necessarily agree on their input bits. In addition, each agent holds a binary *output bit* (also called *opinion*). The goal of all agents is to converge their opinion on the majority bit among the initial input bits of the sources, termed b_{maj}. This problem aims to capture scenarios in which some individuals view themselves as informed, but some of these agents could also be wrong, or not up-to-date. Such situations are common in nature [**CKFL05, REF13**] as well as in man-made systems. The number of sources is termed k. We do not assume that agents know the value k, or that sources know whether they are in the majority or minority (in terms of their input bit). For simplicity, to avoid dealing with the case that the fraction of the majority input bit among sources is arbitrarily close to that of the minority input bit, we shall guarantee convergence only when the fraction of source agents holding the majority input bit is bounded away from $1/2$.

The particular case where we are promised to have $k = 1$ is the (single-source) bit dissemination. In this case we have a single source agent that aims to disseminate its input bit b to the rest of the population, and there are no other sources introducing a conflicting opinion. Note that this problem has been studied extensively in different models under different names (e.g., *broadcast* or *rumor spreading*). Here we use the term *bit dissemination* to focus on the fact that we are interested in the dissemination of a single bit $b \in \{0, 1\}$.

A classical example of bit dissemination considers the synchronous $\mathcal{PUSH}$ /$\mathcal{PULL}$ communication model, where b can be propagated from the source to all other agents in $\mathcal{O}(\log n)$ rounds, by simply letting each uninformed agent copy it whenever it sees an informed agent [**KSSV00**]. The correctness of this protocol heavily relies on the absence of incorrect information in the memory of the agents. Such reliability however may be difficult to achieve in dynamic or unreliable conditions. For example, if the source is sensitive to an unstable environment, it may change its mind several times before stabilizing to its final opinion. Meanwhile, it may have already invoked several consecutive executions of the protocol with contradicting initial opinions, which may in turn "infect" other agents with the wrong opinion $1 - b$. If agents do not share a common time notion, it is unclear how to let infected agents distinguish their current wrong opinion from the more "fresh", correct opinion. To address such difficulty, we consider the context of *self-stabilization* [**Dij74**], where agents must converge to a correct configuration from any initial configuration of states.

2.6.1. Difficulties and intuition on bit dissemination

Consider the bit dissemination problem (where we are guaranteed to have a single source agent). This particular case is already difficult in the self-stabilizing context if we are restricted to use $O(1)$ bits per interaction. As hinted above, a main difficulty lies in the fact that agents do not necessarily share a common time notion. Indeed, it is easy to see that if all agents share the same clock, then convergence can be achieved in $\mathcal{O}(\log n)$ time, with high probability,and using less than three bits per interaction, as described in the following paragraphs.

2.6.1.1. *Solving self-stabilizing bit dissemination ($k = 1$) with 2 bits per interaction, assuming synchronized clocks.* The source sets her output bit to be her input bit b. In addition to communicate her output bit b_u, each agent u stores and communicates a *certainty* bit c_u. Informally, having a certainty bit equal to 1 indicates that the agent is certain of the correctness of its output bit. The source's certainty bit is always set to 1. Whenever a non-source agent v observes u and sees the tuple (b_u, c_u), where $c_u = 1$, it copies the output and certainty bits of u (*i.e.,* sets $b_v = b_u$ and $c_v = 1$). In addition, all non-source agents count rounds, and reset their certainty bit to 0 simultaneously every $T = \mathcal{O}(\log n)$ rounds. The reset allows to get rid of "old" output bits that may result from applying the protocol before the source's output bit has stabilized. This way, from the first time a reset is applied after the source's output bit has stabilized, the correct source's output bit propagates to all agents within T rounds, w.h.p. Note however, that if agents do not share a consistent notion of time they cannot reset their certainty bit to zero simultaneously. In such cases, it is unclear how to prevent agents that have just reset their certainty bit to 0 from being "infected" by "misleading"

agents, namely, those that have the wrong output bit and certainty bit equal to 1.

2.6.1.2. *Solving self-stabilizing bit dissemination ($k = 1$) with a single bit per interaction, assuming synchronized clocks.* Under the assumption that all agents share the same clock, the following trick shows how to obtain convergence in $\mathcal{O}(\log n)$ time and using only a single bit per message, namely, the output bit. As before, the source sets her output bit to be her input bit b. Essentially, agents divide time into phases of some prescribed length $T = \mathcal{O}(\log n)$, each of them being further subdivided into 2 subphases of length $T/2$. In the first subphase of each phase, non-source agents are *sensitive* to opinion 0. This means that whenever they see a 0 in the output bit of another agent, they turn their output bit to 0, but if they see 1 they ignore it. Then, in the second subphase of each phase, they do the opposite, namely they switch their output bit to 1 as soon as they see a 1 (see Figure 9). Consider the first phase starting after initialization. If $b = 0$ then within one complete subphase $[1, T/2]$, every output bit is 0, w.h.p., and remains there forever. Otherwise, if $b = 1$, when all agents go over a subphase $[T/2+1, T]$ all output bits are set to 1, w.h.p., and remain 1 forever. Note that a common time notion is required to achieve correctness.

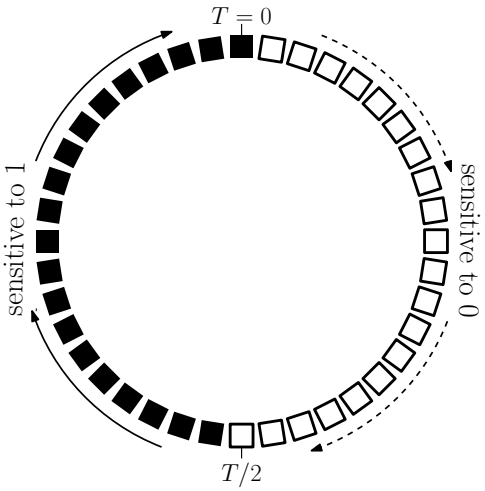

FIGURE 9. The division in subphases used for self-stabilizing bit dissemination with a clock. During the first half, between times 1 and $T/2$, agents are sensitive to 0. Then they are sensitive to 1.

The previous protocol indicates that the self-stabilizing bit dissemination problem is highly related to the self-stabilizing *clock synchronization* problem, where each agent v internally stores a clock modulo $T = T(v) = \mathcal{O}(\log n)$ incremented at every round and, despite having arbitrary initial states (i.e. at the beginning it may be that $T(u) \neq T(v)$ for some $u \neq v$),

all agents should converge on sharing the same value of the clock. Indeed, given such a protocol, one can obtain a self-stabilizing bit dissemination protocol by running the clock synchronization protocol in parallel to the last example protocol. This parallel execution costs only an additional bit to the message size and a $\mathcal{O}(\log n)$ additive factor to the time complexity over the complexities of the clock synchronization protocol.

To synchronize clocks modulo T in a self-stabilizing manner, one could use the stabilizing consensus protocol in [**DGM$^+$11**], by displaying all the bits of the clocks in each message, and reaching consensus on each of them separately and in parallel, while incrementing the clocks (see Section 9.0.3 for further details). Unfortunately, this approach is wasteful in terms of message size, as it requires to reveal $\log T = \mathcal{O}(\log \log n)$ bits per interaction. As another approach, one could aim at sequentially synchronizing clocks bit after bit. That is, first display and synchronize the first bit; then, once agents "know" that the first bit has been synchronized, display and synchronize the second bit, etc. This approach is problematic in the context of self-stabilization, since, first, it requires agents to "know" when a bit is synchronized, and second, it requires agents to agree on the bit index that they currently aim to synchronize. Both of these seem to require clocks to be synchronized to begin with.

2.6.1.3. *Intuition behind the self-stabilizing clock synchronization algorithm.* Our technique for obtaining the clock synchronization protocol is based on a compact recursive use of the stabilizing consensus protocol proposed by Doerr et al. [**DGM$^+$11**] through our Message Reduction Theorem (Theorem 17). In the Section 9.0.3 of Chapter 9, we describe a simple protocol called Syn-Simple that uses $\mathcal{O}(\log T)$ bits per message. In Syn-Simple, each agent u maintains a clock $C_u \in [0, T-1]$. At each round, each agent u displays the opinion of her clock, pulls 2 other such clock opinions, and updates her clock as the bitwise majority of the two clocks she pulled and her own. Then the clock C_u is incremented. This protocol essentially amounts to running the protocol of Doerr et al. on each bit separately and in parallel, and self-stabilizes in $\mathcal{O}(\log T \log n)$ rounds, w.h.p. (Proposition 2).

We want to apply a strategy similar to Syn-Simple, while using only $\mathcal{O}(1)$ many bits per interaction. The core technical ingredient, made rigorous in the Message Reduction Theorem, is that a certain class of protocols using messages of ℓ bits, to which Syn-Simple belongs, can be emulated by another protocol which uses $\lceil \log \ell \rceil + 1$ bits only (see Figure 29). The idea is to build a clock modulo ℓ using Syn-Simple itself on $\lceil \log \ell \rceil$ bits and sequentially display one bit of the original ℓ-bit message according to such clock. Thus, by applying such strategy to Syn-Simple itself, we use a smaller clock modulo $\ell' \ll \ell$ to synchronize a clock modulo ℓ. Iterating such process, in Section 9.2.2, we obtain a compact protocol which uses only 3 bits.

2.6.2. Results of Chapter 9

The main results presented in Chapter 9 are the following.

THEOREM 15 (SYN-PHASE-SPREAD). *Fix an arbitrarily small constant $\varepsilon > 0$. There exists a protocol, called* SYN-PHASE-SPREAD*, which solves the majority bit dissemination problem in a self-stabilizing manner in $\tilde{\mathcal{O}}(\log n)$ rounds*[23]*, w.h.p. using 3-bit messages, provided that the majority bit is supported by at least a fraction $\frac{1}{2} + \varepsilon$ of the source agents.*

Theorem 15 is proved in Section 9.3. The core ingredient of SYN-PHASE-SPREAD is our construction of an efficient self-stabilizing T-clock synchronization protocol, which is used as a black-box. As for the majority bit dissemination problem, the case that interests us is when $T = \tilde{\mathcal{O}}(\log n)$. Note that in this case, the following theorem, proved in Section 9.2, states that the convergence time of the clock synchronization algorithm is $\tilde{\mathcal{O}}(\log n)$.

THEOREM 16 (SYN-CLOCK). *Let T be an integer. There exists a self-stabilizing T-clock synchronization protocol, called* SYN-CLOCK*, which employs only 3-bit messages, and synchronizes clocks modulo T within $\tilde{\mathcal{O}}(\log n \log T)$ rounds, w.h.p.*

The proof of Theorem 16 is given in Section 9.2. In addition to the self-stabilizing context our protocols can tolerate the presence of Byzantine agents. Specifically, it is possible to show that, as a corollary of the analysis given in Chapter 9 and the fault-tolerance property of the analysis in [**DGM+11**], if $T \leqslant poly(n)$ then SYN-CLOCK can tolerate the presence of up to $\mathcal{O}(n^{1/2-\varepsilon})$ Byzantine agents for any $\varepsilon > 0$. In addition, SYN-PHASE-SPREAD can tolerate $\min\{(1-\varepsilon)(k_{maj} - k_{min}), n^{1/2-\varepsilon}\}$ Byzantine agents, where k_{maj} and k_{min} are the number of sources supporting the majority and minority opinions, respectively. Note that for the case of a single source ($k = 1$), no Byzantine agents are allowed; indeed, a single Byzantine agent pretending to be the source with the opposite opinion can clearly ruin any protocol. However, in order to focus on the self-stabilizing aspect of our results, in this work we do not explicitly address the presence of Byzantine agents.

The proofs of both Theorem 16 and Theorem 15 rely on recursively applying a new general compiler which can essentially transform any self-stabilizing algorithm with a certain property (called *bitwise-independence property*) that uses ℓ-bit messages to one that uses only $\lceil \log \ell \rceil + 1$-bit messages, while paying only a small penalty in the running time. This compiler is described in Section 9.1, where we prove the following result. As explained in Section 9.0.2, we denote with $\mathcal{PULL}(\eta, \ell)$ the model in which at each round each node displays ℓ bits in the visible part of her memory, and can observe the visible part of η other agents sampled uniformly at random.

[23] With a slight abuse of notation, with $\tilde{\mathcal{O}}(f(n)g(T))$ we refer to $f(n)g(T) \cdot \log^{\mathcal{O}(1)}(f(n)) \cdot \log^{\mathcal{O}(1)}(g(T))$. All logarithms are in base 2.

THEOREM 17 (Message Reduction Theorem). *Any self-stabilizing protocol Ψ in the $\mathcal{PULL}(\eta, \ell)$ model having the bitwise-independence property, and whose running time is L_Ψ, can be emulated by a protocol* EMUL(Ψ) *which runs in*[24] *the $\mathcal{PULL}(2, \lceil \log(\frac{\eta}{2}\ell) \rceil + 1)$ model, has running time $\mathcal{O}(\log(\eta\ell) \log n + \frac{\eta}{2}\ell L_\Psi)$ and has itself the bitwise-independence property.*

The structure between our different lemmas and results is summarized in the picture below, Figure 26.

As discussed in Chapter 10, it remains an open problem, both for the self-stabilizing bit dissemination problem and for the self-stabilizing clock synchronization problem, whether the message size can be reduced to 2 bits or even to 1 bit, while keeping the running time poly-logarithmic.

[24] The only reason for designing EMUL(Ψ) to run in the

$$\mathcal{PULL}\left(2, \lceil \log\left(\frac{\eta}{2}\ell\right) \rceil + 1\right)$$

model in the Message Reduction Theorem is the consensus protocol we adopt, 3-Median dynamics, which works in the $\mathcal{PULL}(2)$ model. In fact, EMUL(Ψ) can be adapted to run in the

$$\mathcal{PULL}\left(1, \lceil \log\left(\eta\ell\right) \rceil + 1\right)$$

model by using a consensus protocol which works in the $\mathcal{PULL}(1)$ model. However, no self-stabilizing binary consensus protocol in the $\mathcal{PULL}(1)$ model with the same performances as 3-Median dynamics is currently known.

CHAPTER 3

Work Related to Dynamics (and Surroundings)

In this section we aim to provide the scientific context the study of dynamics belongs to, focusing on those dynamics studied in this work, and on classes of protocols closely related to them.

3.1. Dynamics for Community Detection

Dynamics have received considerable attention across different research communities, both as efficient distributed algorithms [**AAE08, BTV09, OT09, MRSDZ11**] and as abstract models of *natural* interaction mechanisms inducing emergent behavior in complex systems [**AAB+11, CCN12, Dot14, FPM+02, MNT14**]. For instance, simple averaging dynamics have been considered to model opinion formation mechanisms [**DeG74, FJ90**], while a number of other dynamics have been proposed to describe different social phenomena [**EK10**].

An important class of protocols which includes a wide range of dynamics is that of *label propagation algorithms*.

3.1.1. Label Propagation Algorithms

Label propagation algorithms (LPA for short) [**RAK07**] are a class of protocols based on a simple epidemic mechanism which can be efficiently implemented in a fully-distributed fashion, since they require easy local computations. In their most famous basic version, some distinct labels are initially assigned to a subset of nodes; at every step, each node updates her label (if any) by choosing the label which most of her (current) neighbors have (the *majority* label); if there are multiple majority labels, one label is chosen randomly. Typically, the goal of the protocol is to converge to a good labeling which reflects the clustered structure of the graph.

We remark that while the (informal) notion of LPA does suggest a more restricted set of possible update functions for the nodes' states compared to the general notion of dynamics, an LPA algorithm is not necessarily a dynamics: for example, an LPA is not necessarily time-homogeneous [**CDIG+15**]. However, a protocol which is both a dynamics and an LPA is an ideal representative of both classes of protocols, such as the 3-Majority dynamics analyzed in Chapter 5.

Despite the simplicity of LPA-based protocols, very few analytical results are known on their performance over relevant classes of graphs. It seems hard

to derive, from empirical results, any fundamental conclusions about LPA behavior, even on specific families of graphs [**KPS13**]. One reason for this hardness is that despite its simplicity, even on simple graphs, the class of LPA can exhibit complex behavior, not far from epidemic processes such as the spread of a disease in an interacting population [**New02**].

Several versions of LPA-based protocols have been tested on a wide range of social networks [**RAK07, BC09, LHLC09, LM10, CG12**]: such works experimentally show that LPA-based protocols work quite efficiently and are effective in providing *almost* good labeling. Based on extensive simulations, Raghavan et al. [**RAK07**] and Leung et al. [**LHLC09**] empirically show that the average convergence time of LPA-based protocols is bounded by some logarithmic function of n on special classes of graphs whose community structure is known.

The only available rigorous analysis of label propagation algorithms on the stochastic block model $\mathcal{G}_{2n,p,q}$ is the one presented in [**KPS13**], where the authors propose and analyze a label propagation protocol on $\mathcal{G}_{2n,p,q}$ for highly-dense topologies. In particular, their analysis considers the case where $p = \Omega(1/n^{1/4-\varepsilon})$ and $q = \mathcal{O}(p^2)$, a parameter range in which very dense clusters of constant diameter separated by a sparse cut occur w.h.p. In this setting, characterized by a polynomial gap between p and q, simple combinatorial and concentration arguments show that the protocol converges in constant expected time. They also conjecture a logarithmic bound for sparser topologies.

In general, providing analytical bounds on the convergence time of LPA-based protocols over relevant classes of networks is an important open question that has been proposed in several papers arising from different areas [**RAK07, LHLC09, BC09, CG12, KPS13, KMTN15**]. The results about the 3-Majority and Undecided-State dynamics studied in Chapter 5 and Chapter 6, and the related work discussed in sections 3.4 and 3.3, represent some preliminary contributions with this respect.

Before moving to discuss popular solutions for the community detection problem, we review one of the most popular random graph models which have been theoretically investigated in order to understand the average case complexity of community detection.

3.1.2. Stochastic block models for average case community detection

Probably the most natural way to formalize a basic instance of the community detection problem is as an instance of the minimum bisection problem: Given a graph $G = (V, E)$ with $|V| = 2n$, we are asked to find the partition (bisection) $V_1, V_2 \subset V$ with $V_1 \dot{\cup} V_2 = V$ such that $|\{(u, v) \in E : u \in V_1, v \in V_2\}|$ is minimized. Unluckily, the minimum bisection problem was one of the first problems to be shown NP-complete

[**GJS76**]. Therefore, when complexity theorists realized that, despite a problem being NP-hard *in the worst case*, it is still possible to get precious insights on its structure by investigating its *average case complexity*, the problem of community detection did not wait much until a natural formulation to investigate its average-case structure was proposed [**DF89**].

The *stochastic block model* is arguably the simplest random graph model which exhibits a community structure, and in its basic form it can be described as two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ of n nodes (the *communities*) generated according to an Erdős-Rényi model with parameter $p = a/n$, which are "glued" together by adding an edge (u, v) between each pair of nodes $u \in V_1$ and $v \in V_2$ with probability $q = b/n \leqslant p$ (see Definition 6) [**HLL83**]. Observe that if $q = p$ we have an Erdős-Rényi model with parameter p over $2n$ nodes, while the cut between the two communities becomes intuitively much easier to detect whenever $q \ll p$.

Because of their naturalness, stochastic block models have been deeply studied in statistics [**HLL83, MNS14**], computer science [**Bop87, DF89, Mas14**], probability theory [**MNS14**], statistical physics [**DKMZ11, KMM$^+$13**], and social sciences [**HLL83**].

In the connected regime with $a = \Omega(\log n)$, the communities can be exactly recovered and a sharp exact recovery threshold is known [**ABH14**]. Exact recovery thresholds have also been identified in a more general setting with a fixed number of communities, and with heterogeneous community sizes and link probabilities [**ABH14**]. However, real networks are often sparse with bounded average degrees, and in the sparse setting with $a = o(\log n)$ exact recovery of communities from the graph becomes impossible [**CO05**]. Thus the goal in the sparse regime is to find a labeling that has good correlation with the true one (up to permutation of community labels) [**KMM$^+$13**].

As discussed in Section 3.1.4, efficient algorithms for stochastic block models were developed and shown to detect the blocks whenever this is theoretically possible. Finally, we remark that it is known that no local algorithm with access to neighborhoods of radius $o(\log n)$ (thus any distributed algorithm in the $\mathcal{LOCAL}$ model operating in less than $\log n$ rounds), can have non-trivial performance for this problem [**GS14**].

Because of their relevance for the reconstruction problem, in the next section we discuss a class of algorithms, *belief propagation algorithms*, whose simplicity is close to that of dynamics.

3.1.3. Belief propagation algorithms

Belief propagation algorithms are best known as message-passing algorithms for performing inference in graphical models [**Mac03**]. Belief propagation cannot be considered a dynamics: At each round, each node sends a different message to each neighbor, which means that the update rule is not symmetric w.r.t. the neighbors, thus requiring port numbering [**Suo13**],

and the required local memory grows linearly in the degree of the node. Non-rigorous methods have given strong evidence that some *belief propagation algorithms* are optimal for the reconstruction problem [**DKMZ11**]. Their rigorous analysis is a major challenge; in particular, the convergence to the correct value of belief propagation is far from being fully-understood on graphs which are not trees [**Wei00, MK07**].

As we discuss in the next subsection, more complex algorithms, many of which have been inspired by belief propagation, have been rigorously shown to perform reconstruction optimally.

3.1.4. General algorithms for the reconstruction problem

While improving performance of spectral clustering algorithms and testing their limits for the purpose of the reconstruction problem is not the main driver behind our study in Chapter 4, for the sake of completeness, we next compare our results on the Averaging dynamics to the previous general algorithms for the reconstruction problem.

Several algorithms for community detection are *spectral*: They typically consider the eigenvector associated to the second eigenvalue of the adjacency matrix A of G, or the eigenvector corresponding to the largest eigenvalue of the matrix $A - \frac{d}{n}J$ [**Bop87, CO05, CO10, McS01**][1], on the grounds that these are correlated with the hidden partition. In [**AS15, CO10, MNS13, KMM$^+$13, BLM15**], spectral algorithms have been proposed that find a weak reconstruction even in the sparse, tight regime of the stochastic block model, where an eigenvalue computation can be used to find an approximation of the hidden partition which, in certain cases, can be refined to an exact computation of the hidden partition using a post-processing phase.

Even though the above mentioned algorithms have been presented in a centralized setting, spectral algorithms turn out to be a feasible approach also for distributed models. Indeed, Kempe and McSherry [**KM04**] show that eigenvalue computations can be performed in a distributed fashion, yielding distributed algorithms for community detection in various models, including the stochastic block model. However, Kempe and McSherry's algorithm as well as any distributed version of the above mentioned centralized algorithms are not dynamics. Actually, adopting the effective concept from Hassin and Peleg in [**HP01**], such algorithms are also not *light-weight*: Different and not-simple operations are executed at different rounds, nodes have identities, messages are treated differently depending on the originator, and so on. Moreover, a crucial aspect is convergence time: The mixing time of the simple random walk on the graph is a bottleneck for Kempe and McSherry's algorithm and for any algorithm that performs community detection in a graph G, by employing the power method or the Lanczos method [**Lan50**] as a subroutine to compute the eigenvector associated to the second eigenvalue

[1] A is the adjacency matrix of G, J is the matrix having all entries equal to 1, d is the average degree and n is the number of nodes.

of the adjacency matrix of G. Notice that the mixing time of graphs sampled from $\mathcal{G}_{2n,p,q}$ is concentrated around $\frac{a+b}{2b}$: hence, it can be super-logarithmic and even $n^{\Omega(1)}$.

In general, the reconstruction problem on the stochastic block model has been studied extensively using a multiplicity of techniques, which include combinatorial algorithms [**DF89**], belief propagation [**DKMZ11**], spectral-based techniques [**McS01, CO10**], Metropolis approaches [**JS98**], and semi-definite programming [**ABH14**], among others. Unlike the distributed setting, where the existence of *light-weight protocols* [**HP01**] is the main issue (even in non-sparse regimes), in centralized setting strong attention has been devoted to establishing sharp thresholds for weak and strong reconstruction. Define $a = np$ as the expected *internal degree* (the number of neighbors that each node has on the same side of the partition) and $b = nq$ as the expected *external degree* (the number of neighbors that each node has on the opposite side of the partition). Decelle et al. [**DKMZ11**] conjectured that weak reconstruction is possible if and only if $a - b > 2\sqrt{a+b}$. This was proved by Massoulie and Mossel et al. [**MNS13, Mas14, MNS14**]. Strong recovery is instead possible if and only if $a - b > 2\sqrt{a+b} + \log n$ [**ABH14**].

Versions of the stochastic block model in which the random graph is regular have also been considered [**MNS14, BDG$^+$16**]. In particular Brito et al. [**BDG$^+$16**] show that strong reconstruction is possible in polynomial-time when $a - b > 2\sqrt{a+b-1}$.

In the next section we review the literature concerning the Averaging dynamics, which is the main ingredient of our protocol in Chapter 4.

3.2. The Averaging Dynamics

The Averaging dynamics, in which each node updates its value to the average of its neighbors, is perhaps one of the simplest and most interesting examples of linear dynamics and it always converges when G is connected and not bipartite: It converges to the global average of the initial values if the graph is regular and to a weighted global average if it isn't [**BGPS06, Sha09**]. Important applications of linear dynamics have been proposed in [**KDG03, AYSS09, Tsi84, Kle99**], for example to perform basic tasks such as self-stabilizing *consensus* in faulty distributed systems [**BTV09, XBK07, OT09**]. The convergence time of the Averaging dynamics is the mixing time of a random walk on G [**Sha09**]. It is logarithmic in $|V|$ if the underlying graph is a good *expander* [**HLW06**], while it is slower on graphs that exhibit sparse cuts.

While the Averaging dynamics is based on the statistical concept of *average*, in the next section we discuss the previous work related to designing a dynamics based on the statistical concept of *mode*.

3.3. Dynamics for Plurality Consensus

In this section we discuss previous work related to the results of chapters 5 and 6 regarding the plurality consensus problem. The plurality consensus problem arises in several applications such as distributed database management, where data redundancy or replication and majority rules are used to manage the presence of unknown faulty processors [**DGM$^+$11, Pel02**]. The goal here is to converge to the version of the data supported by the plurality of the initial distributed copies (it is reasonable that a sufficiently strong plurality of the nodes are not faulty and thus possess the correct data). Another application is distributed item ranking, in which every node initially selects some item and the goal is to agree on the most popular item according to the initial plurality opinion [**PVV09**]. Further applications of majority updating rules in networks can be found in [**EK10, Pel02**].

Results closely related to those in Chapter 5 are those in [**DGM$^+$11**]. Several variants of binary majority consensus have been studied in different distributed models [**AAE08, MS10**]. The simplest protocol is the polling rule, i.e. the 1-majority dynamics, which has been extensively studied on several classes of graphs (see [**Pel02**]).

As for the *population model*, where there is only one random node-pair interaction per round (so the dynamics are strictly sequential), the binary case on the clique has been studied in [**AAE08**] where the *Undecided-State* dynamics has been introduced. Their generalization to the multivalued case ($k \geqslant 3$) does not converge to plurality even starting with a large bias $s = \Theta(n)$. Following [**AAE08**], [**MNRS14**] has analyzed a similar protocol on general graphs which solves the binary majority consensus deterministically. In [**AGV15**], the trade-off between deterministic success and convergence speed for protocols solving the binary majority consensus problem in population protocols is investigated.

More expensive and complex protocols have been considered in order to speed up the process. For instance, in [**KT08**], a protocol for the sequential-interaction model is presented that requires $\Theta(\log n)$ memory per node and converges in time $\mathcal{O}(n^7)$. Other protocols for the sequential-interaction model have been analyzed in [**BTV09, LB95**] (with no time bound).

In [**PVV09, DV12, BD13**], the Undecided-State dynamics on the continuous-time population model is proved to converge in $\mathcal{O}(n \log n)$ expected time only for $k = \Theta(1)$ and $s = \Theta(n)$: Even assuming such strong restrictions, the bound does not hold with high probability and, moreover, their analysis, based on real-valued differential-equations, do not work for the discrete-time parallel model considered in Chapter 5.

Protocols for specific network topologies and some "social-based" communities have been studied in [**AD15, DV12, MNT14, PVV09**]. We mention that similar majority-consensus problems have also been studied (for example in [**AD15, MNT14**]) in the *$\mathcal{LOCAL}$ (communication) model*

[**FKP11, Pel00**] where, however, node congestion and memory size are linear in the node degree of the network.

In [**KDG03**], the authors provide a protocol in the uniform $\mathcal{PUSH}$ model to compute aggregate functions, which can be used to solve plurality consensus in polylog(n) time starting from any positive bias, but it requires exponentially larger memory and message size than the 3-Majority dynamics and Undecided-State dynamics (namely $\Theta(k \log n)$). Moreover, their protocol requires the nodes to send slightly more complex messages than their sole current opinion, and its effectiveness heavily relies on a potential function argument that is sensible to slight changes to the model (e.g. it does not hold in the presence of noise, which is the variant of the plurality consensus problem which we consider in Section 2.5).

Finally, in [**CER14**], the authors provide a rigorous analysis of a simple 2-voting dynamics for the binary case on any (possibly random) regular graph: in the latter case, they provide optimal bounds on the convergence time as a function of the second-largest eigenvalue of the graph.

The major aforementioned contributions to the plurality consensus problem, prior to the analysis of the Undecided-State dynamics provided in Chapter 6, are summarized in Figure 10.

In the next section we briefly review the previous work regarding the Undecided-State dynamics for plurality consensus, which is the main character of Chapter 6.

3.4. Undecided-State Dynamics

The Undecided-State dynamics has been introduced and analyzed in [**AAE08**] for the binary case in the population protocol model (where only one edge is active during a round). They prove that this dynamics has "parallel" convergence time $O(\log n)$ whenever the bias $\Omega\left(\sqrt{n \log n}\right)$. The same dynamics has been analyzed in different distributed models for the binary case [**BD13, BTV09, DV12, PVV09, MRSDZ11**], or when k is an absolute constant [**JKV12**]. Last but not least, interest for this dynamics has been stimulated by findings in biology: notably, as shown in [**CCN12**], the structure and dynamics of the "approximate majority" protocol (as it is called there and in [**AAE08**]) is to a great extent similar to a mechanism that is collectively implemented in the network that regulates the mitotic entry of the cell cycle in eukaryotes.

In the next section we leave behind the requirement of converging to an initial value with some property (i.e. being the plurality), and we review the literature concerning the problem of converging to any initial value and maintaining consensus on that value even in the presence of an adversary.

	Mem. & mess. size	**# of colors**	**Time efficiency**	**Comm. Model**
Kempe et al. FOCS '03	$O(k \log n)$	any	$O(\log n)$	$\mathcal{PUSH}$
Angluin et al. DISC '07 Perron et al. INFOCOM '09	$\Theta(1)$	2	$O(\log n)$	Sequential
Doerr et al. SPAA '11	$\Theta(1)$	2	$O(\log n)$	$\mathcal{PULL}$
Babaee et al. Comp. J. '12 Jung et al. ISIT '12	$O(\log k)$	Constant	$O(\log n)$	Sequential
Becchetti et al. SPAA '14	$O(\log k)$	$n^{\Theta(1)}$	$O(k \log n)$	$\mathcal{PULL}$

FIGURE 10. The table summarizes the major previous contributions toward an efficient dynamics for plurality consensus in random interaction models. The time efficiency in the last row is that of the 3-Majority dynamics, which motivated the research that led to the Undecided-State dynamics.

3.5. Dynamics for Stabilizing Consensus

Consensus problems in distributed systems have been the focus of a large body of work in several research areas, such as distributed computing [**GK10**], communication networks [**RM08**], social networks and voting systems [**MNT14, YOA+13**], distributed databases [**DGH+87, DGM+10**], biological systems and chemical reaction networks [**CCN12**]. For brevity's sake, we here focus on results that are closest in spirit to the results of Chapter 5 regarding the stabilizing consensus problem. (Part of the literature has already been mentioned in the previous section regarding plurality consensus: we mention some of those works again to discuss them in a different perspective and to make the section self-contained.)

In [**AAE08**], the authors show that n agents that meet at random can reach valid stabilizing almost-consensus in $\mathcal{O}(n \log n)$ pairwise interactions, w.h.p., even against an $F = o(\sqrt{n})$-bounded dynamic adversary. The adopted protocol is the Undecided-State protocol [**AAE08, PVV09**],

discussed in sections 3.3 and 3.4. However, their analysis (and, thus, their result) only holds for the binary case and for the *population-protocol* model: At every round only one pair of nodes can interact. The authors left the existence of protocols for the multi-valued Byzantine case as a final open question [**AAE08**].

In the uniform $\mathcal{PULL}$ model, in [**DGM+11**] the authors provide an analysis of the *3-Median* dynamics, in which every node updates her value to the median of her random sample. They show that this dynamics converges to an almost-agreement configuration (which is even a good approximation of the global median) within $\mathcal{O}(\log k \cdot \log\log n + \log n)$ rounds, w.h.p. It turns out that, in the binary case, the median rule is equivalent to the 3-Majority dynamics, thus their result implies that the 3-Majority dynamics is an ($F = \sqrt{n}$-)stabilizing consensus with $\mathcal{O}(\log n)$ convergence time. However, in the non-binary case, it requires Σ to be a totally-ordered set and this order to be *consistent*, i.e. all agents agree on it: This may be a strong restriction when these processes are used to model emerging behavior and self-organization in complex agent systems such as biological ones.

Unfortunately, even assuming an ordered opinion set $(\Sigma, \leqslant)$, the 3-Median dynamics does not guarantee the crucial property of *validity* against both F-static (and, clearly, dynamic) adversaries for small bounds on F (see Figure 11). The latter deficiency of the 3-Median dynamics is critical, since

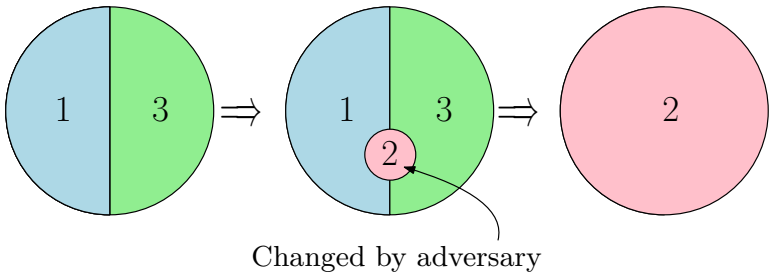

FIGURE 11. It is not hard to see that from the configuration with $n/2$ agents holding value 1 and $n/2$ agents holding value 2, an adversary with power roughly $\sqrt{n}$ can lead the system to converge to value 2 which is not initially present in the system, and thus not valid.

the validity property of consensus plays a crucial role in several realistic scenarios, such as monitoring sensor networks, bio-inspired dynamic systems, and voting systems [**CCN12, MNT14, YOA+13**]. Another version of binary stabilizing almost-consensus is the one studied in [**YOA+13**]: Here, corrupted nodes are *stubborn* agents of a social network who influence others but never change their opinions. They prove negative results under a generalized variant of the classic polling dynamics [**HP01**] in the (Poisson-clock) population-protocol model.

In the next sections, we move from work directly linked to dynamics to the broader literature pertinent to chapters 7, 8 and 9.

3.6. Repeated Balls-into-Bins and Random Walks in the Uniform $\mathcal{PUSH}$ Model

In this section we briefly review the literature concerning the repeated balls-into-bins process considered in Chapter 7. Recall that the repeated balls-into-bins process is equivalent to the process of performing parallel random walks in the (uniform) $\mathcal{GOSSIP}$ model, and in Chapter 7 we investigate the maximum load of the former in order to bound the congestion of the latter.

3.6.0.1. *Random walks on graphs.* The original process of parallel random walks in the (uniform) $\mathcal{GOSSIP}$ model (also known as random phone-call model [**DGH$^+$87, KSSV00**]), was first considered in [**BCEG10, BCN$^+$15a, EK15**], when every message can contain at most one token. Maximum load (i.e., node congestion), token delays, mixing and cover times are here the most crucial aspects. We remark that the flavor of these studies is different from that of Chapter 7: indeed, their main goal is to keep maximum load and token delays logarithmic over some *polylogarithmic period*. Their aim is to achieve a fast mixing time for every random walk in the case of good expander graphs. In particular, in [**BCEG10**], a logarithmic bound is shown for the complete graph when $m = \mathcal{O}(n/\log n)$ random walks are performed over a logarithmic time interval. A similar bound is also given for some families of almost-regular random graphs in [**EK15**].

3.6.0.2. *Parallel computing.* Balls-into-bins processes have been extensively studied in the area of parallel and distributed computing, mainly to address balanced-allocation problems [**ABKU99, BCSV06, RS98**], PRAM simulation [**KLMadH96**] and hashing [**DGM$^+$10**]. In order to optimize the total number of random bin choices used for the allocation, further allocation strategies have been proposed and analyzed (see, e.g., [**ACMR95, BKSS13, Mit01, MPS02, Vöc03**]). As mentioned in Section 2.4, the notion of stability adopted in chapters 5 and 7 is inspired by those investigated in [**AKU05, BFG03, BFK$^+$16**] where load balancing algorithms are analyzed in scenarios in which new tasks arrive during the evolution of the system, and existing jobs are executed by the processors and leave the system. An adversarial model for a sequential balls-into-bins process has been studied in [**AS09**]. We remark that, in the above previous works, the goal is different from ours: each ball/task must be allocated to *one, arbitrary* bin/processor (it is not a token-traversal process).

3.6.0.3. *Queuing theory.* In classical queuing theory the closest model to the setting considered in Chapter 7 is the *closed Jackson network* [**Asm03**]. In this model, time is continuous and each node processes a single token among those in its queue; processing each token takes an exponentially distributed

interval of time. As soon as its processing is completed, each token leaves the current node and enters the queue of a neighbor chosen uniformly at random. Notice that, since time is continuous, the process' events are sequential, so that the associated Markov chain is much simpler than the one describing our parallel process. In particular, the stationary distribution of a closed Jackson network can be expressed as a product-form distribution. It is noted in [**HW92**] that "[...] virtually all of the models that have been successfully analyzed in classical queuing network theory are models having a so-called product form stationary distribution". Given the non-reversibility of the Markov chain associated to the repeated balls-into-bins process and other difficulties discussed in Chapter 7, the stationary distribution is instead very likely not to exhibit a product-form distribution, thus laying outside the domain where the techniques of classical queuing theory seem effective.

3.6.0.4. *Queuing systems in computer science.* Among the works in computer science which depart from the classical framework of queueing theory, we remark the seminal work [**BKR$^+$01**] on *adversarial queuing systems*: here, new tokens (having specified source and destination nodes) are inserted in the nodes according to some adversarial strategy and a notion of *edge-congestion* stability is investigated. We also note that a probabilistic version of the TETRIS process (which we investigate in Chapter 7 in order to bound the congestion of parallel random walks in the $\mathcal{PUSH}$ model), has been studied in [**BFGK16**]. There, the number of new balls arriving at each round is a random variable with expectation λn, for some $\lambda = \lambda(n) \in [0, 1]$,

In the next, final section, we review the literature about "biological distributed algorithms" pertinent to Chapter 9 (and, partly, to Chapter 8). Thus, the next section focus especially on the bit dissemination and plurality consensus problems.

3.7. Toward a Dynamics for Self-Stabilizing Bit Dissemination

The computational study of abstract systems composed of simple individuals that interact using highly restricted and stochastic interactions has been gaining considerable attention in the community of theoretical computer science. Popular models include *population protocols* [**AAD$^+$06, AR07, AAFJ08, BBK11**], which typically consider constant size individuals that interact in pairs (using constant size messages) in random communication patterns, and the *beeping* model [**AAB$^+$11, EW13**], which assumes a fixed network with extremely restricted communication. The models considered in chapters 8 and 9 also falls in this framework as we consider the uniform $\mathcal{PUSH}$ and $\mathcal{PULL}$ models [**DGH$^+$88, KSSV00, KDG03**] with constant size messages. So far, despite interesting works that consider different fault-tolerant contexts [**AAE08, AAFJ08, BBK11**], most of the progress in this framework considered non-faulty scenarios.

Information dissemination is one of the most well-studied topics in distributed computing, see *e.g.* [**AAE08, DGM+11, DGH+88, CHHKM12, DF11, FHK14, KSSV00**]. Classical examples include the *bit dissemination* (*broadcast* or *rumor-spreading*) problem, in which a piece of information residing at one source agent is to be disseminated to the rest of the population, and *majority consensus* problems in which processors are required to agree on a common output value which is the majority initial input value among all agents [**AAE08, KK13**], or among a set of designated source agents [**FHK14**]. An extensive amount of research has been dedicated to study such problems in the $\mathcal{PUSH}/\mathcal{PULL}$ communication models (including the *phone call* model), due to the inherent simplicity and fault-tolerant resilience of such meeting patterns. Indeed, the robustness of $\mathcal{PUSH}/\mathcal{PULL}$ based protocols to weak types of faults, such as crashes of messages and/or agents, or to the presence of relatively few Byzantine agents, has been known for quite a while [**ES09, KSSV00**]. In [**FHK14**], it has been shown that under the $\mathcal{PUSH}$ model, there exist efficient bit dissemination and majority consensus protocols that use a single bit per message and can overcome flips in messages (noise). The protocols therein, however, assume that the messages are binary and heavily rely on the assumption that agents know when the protocol has started. Observe that in a self-stabilizing context, in which the adversary can corrupt the initial clocks setting them to arbitrary times, such an assumption would be difficult to remove while preserving the small message size.

In general, there are only few known self-stabilizing protocols that operate efficiently under stochastic and capacity restricted interactions. An example is the work of Doerr et al. on *stabilizing consensus* [**DGM+11**] operating in the $\mathcal{PULL}$ model. In that work, each agent initially has a state taken out of a set of m opinions and the goal is to converge on one of the proposed states. The proposed dynamics which runs in logarithmic time is based on sampling the states of 2 agents and updating the agent's state to be the median of the 2 sampled states and the current state of the agent (3 opinions in total). Since the total number of possible states is m, the number of bits that must be revealed in each interaction is $\Omega(\log m)$. Another example is the 3-Majority dynamics studied in Chapter 5, in which each agent has initially an opinion and we want the system to converge to the most frequent one in the initial configuration of the system. In fact, the majority bit dissemination problem studied in Chapter 9 can be viewed as a generalization of the *majority-consensus* problem (i.e. the plurality consensus problem with two opinions) to the case in which multiple agents may initially be without opinion.

Another fundamental issue in distributed computing is *clock synchronization* [**AHR96, Lam78, LLW10, LLSW10**]. We consider a synchronous system in which clocks tick at the same pace but they may not share the same value. This version has earlier been studied in e.g., [**BDH08, Dol97, DH07, DW04, Her00, FK15**] under different names, including

"digital clock synchronization" and "synchronization of phase-clocks"; We simply use the term "clock synchronization". There is by now a substantial line of work on clock synchronization problems in a self-stabilizing context [**Spr13, DW04, LRS15, LR15**]. We note that in these papers the main focus is on the resilience to Byzantine agents. The number of rounds and message lengths are also minimized, but typically as a function of the number of Byzantine processors. The focus of Chapter 9 is instead on minimizing the time and message complexities as much as possible. The authors in [**LRS15, LR15**] consider mostly a deterministic setting, where every agent gets one message from every other agent on each round. Moreover, agents are assumed to have unique identifiers. In contrast, Chapter 9 investigates the restricted and randomized uniform $\mathcal{PULL}$ model. In [**Spr13, LRS15**] randomized protocols are also investigated. We remark that the first protocol we discuss Syn-Simple (Proposition 2), which relies on a known simple connection between consensus and counting [**Spr13**], already improves exponentially on the randomized algorithms from [**Spr13, LRS15**] in terms of number of rounds, number of memory states, message length and total amount of communication, in the restricted regime where the resilience parameter f satisfies $\log n \leqslant f \leqslant n^{\frac{1}{2}-\varepsilon}$ for an arbitrarily small constant $\varepsilon > 0$. We further note that the works [**LRS15, LR15**] also use a recursive construction for their clocks (although very different from the one we use in the proof of Theorem 16). The induction in [**LRS15**] is on the resilience parameter f, the number of agents and the clock length together. This idea is improved in [**LR15**] to achieve optimality in terms of resilience to Byzantine agents.

Finally, we remark that Chapter 9 presents the first analysis investigating the self-stabilizing clock synchronization and majority bit dissemination problem which aims at minimizing the message size beyond logarithmic.

CHAPTER 4

Averaging Dynamics

In this chapter we formally prove the results presented in Section 2.1 on the Averaging dynamics. Recall that in the Averaging protocol, given an underlying graph, initially each node locally chooses a value in $\{-1, 1\}$, uniformly at random and independently of other nodes; Then, in each consecutive round, every node updates her local value to the average of the values held by her neighbors, at the same time applying an elementary, local clustering rule that only depends on the current and the previous values held by the node (Algorithm 1).

As discussed in Section 2.1, while previous work on applications of linear dynamics has focused on tasks that are specific to distributed computing such as reaching consensus or stability in the presence of faulty nodes (see Section 3.2), in this chapter we prove that the process resulting from the Averaging dynamics produces a clustering that exactly or approximately (depending on the graph) reflects the underlying cut in logarithmic time, under various graph models that exhibit a sparse balanced cut, including the stochastic block model. We also prove that a natural extension of this dynamics performs community detection on a regularized version of the stochastic block model with multiple communities.

4.1. Linear Algebra Toolkit

We start by recalling some basic facts from linear algebra and some results from matrix perturbation theory [**SS90**].

If $M \in \mathbb{R}^{n\times n}$ is a real symmetric matrix, then it has n real eigenvalues (counted with repetitions), $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$, and we can find a corresponding collection of orthonormal real eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ such that $M\mathbf{v}_i = \lambda_i \mathbf{v}_i$. If $\mathbf{x} \in \mathbb{R}^n$ is any vector, then we can write it as a linear combination $\mathbf{x} = \sum_i \alpha_i \mathbf{v}_i$ of eigenvectors, where the coefficients of the linear combination are $\alpha_i = \langle \mathbf{x}, \mathbf{v}_i \rangle$. In this notation, we can see that

$$M\mathbf{x} = \sum_i \lambda_i \alpha_i \mathbf{v}_i, \quad \text{and so} \quad M^t\mathbf{x} = \sum_i \lambda_i^t \alpha_i \mathbf{v}_i.$$

Unless otherwise specified, the norm of a vector $\mathbf{x}$ is the ℓ_2 norm $\|\mathbf{x}\| := \sqrt{\sum_i (\mathbf{x}(i))^2}$ and the norm of a matrix A is the spectral norm $\|A\| := \sup_{\mathbf{x}:\|\mathbf{x}\|=1} \|A\mathbf{x}\|$. For a diagonal matrix, this is the largest diagonal entry in absolute value. In the following we recall the Cauchy-Schwarz inequality,

some properties of the ℓ_2 norm, and a matrix version of the Chernoff bound for random matrices.

LEMMA 1 (Cauchy-Schwarz inequality). *For any pair of vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *it holds*

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leqslant \|\mathbf{x}\| \cdot \|\mathbf{y}\|.$$

OBSERVATION 1. For any matrix A and any vector $\mathbf{x}$, it holds

$$\|A\mathbf{x}\| \leqslant \|A\| \cdot \|\mathbf{x}\|, \quad \text{and} \quad \|A \cdot B\| \leqslant \|A\| \cdot \|B\|.$$

THEOREM 18 (Matrix Bernstein Inequality). *Let* $X_1, \ldots, X_N$ *be a sequence of independent* $n \times n$ *symmetric random matrices*[1]*, such that* $\mathbb{E}[X_i] = \mathbf{0}$ *for every* i*, and such that* $\|X_i\| \leqslant L$ *with probability 1 for every* L*. Call* $\sigma := \|\mathbb{E}\sum_i X_i^2\|$*. Then, for every* t*, we have*

$$\Pr\left(\left\|\sum_i X_i\right\| \geqslant t\right) \leqslant 2ne^{\frac{-t^2}{2\sigma + \frac{2}{3}Lt}}.$$

The following theorems are a weaker version than the original ones they are named after. For a proof of the following one, see Corollary 4.10 in [**SS90**].

THEOREM 19 (Weyl's Theorem). *Let* M_1 *and* M_2 *be two Hermitian matrices, let* $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ *be the eigenvalues of* M_1 *with multiplicities in non-increasing order, and let* $\lambda'_1 \geqslant \lambda'_2 \geqslant \cdots \geqslant \lambda'_n$ *be the eigenvalues of* M_2 *with multiplicities in non-increasing order. Then, for every* i*,*

$$|\lambda_i - \lambda'_i| \leqslant \|M_1 - M_2\|.$$

The (general version of the) following theorem was originally proved in [**DK70**].

THEOREM 20 (Davis and Kahan, 1970). *Let* M_1 *and* M_2 *be two symmetric real matrices, let* $\mathbf{x}$ *be a unit length eigenvector of* M_1 *of eigenvalue* t*, and let* $\mathbf{x}_p$ *be the projection of* $\mathbf{x}$ *on the eigenspace of the eigenvectors of* M_2 *corresponding to eigenvalues* $\leqslant t - \delta$*. Then*

$$\|\mathbf{x}_p\| \leqslant \frac{2}{\delta\pi}\|M_1 - M_2\|.$$

4.2. Distributed Reconstruction Problem

Let us recall the definition of strong and weak reconstruction.

[1]We remark that here we are only assuming that, for each $w, z \in \{1, \ldots, N\}$ with $w \neq z$ and $i_w, j_w, i_z, j_z \in [n]$, $(X_w)_{i_w, j_w}$ and $(X_z)_{i_z, j_z}$ are independent. For any $w \in \{1, \ldots, N\}$, no other assumption on the distribution of the entries of X_w is made, as long as for each $i, j \in [n]$ it holds $(X_w)_{i,j} = (X_w)_{j,i}$ with probability 1.

DEFINITION 2 (Strong and Weak Reconstruction). Given a graph $G = (V_1 \cup V_2, E)$ with $V_1 \cap V_2 = \emptyset$, a *weak (block) reconstruction* is a two-coloring of the nodes that separates V_1 and V_2 up to a small fraction of the nodes. Formally, we define an *ε-weak reconstruction* as a map

$$f : V_1 \cup V_2 \to \{\texttt{red}, \texttt{blue}\}$$

such that there are two subsets $W_1 \subseteq V_1$ and $W_2 \subseteq V_2$ with[2]

$$|W_1 \cup W_2| \geqslant (1-\varepsilon)|V_1 \cup V_2| \quad \text{and} \quad f(W_1) \cap f(W_2) = \emptyset.$$

When $\varepsilon = 0$ we say that f is a *strong reconstruction*.

Given a graph $G = ((V_1, V_2), E)$, the reconstruction problem requires computing an ε-reconstruction of G. To this purpose, in this chapter we analyse the distributed protocol given in Algorithm 1 (see also figures 13 and 14), which is based on the Averaging dynamics and produces a coloring of the nodes at the end of every round.

4.2.1. The Averaging dynamics and random walks on G

The analysis of the Averaging dynamics on a graph G is closely related to the behavior of random walks in G, which are best studied using tools from linear algebra that we briefly summarize below.

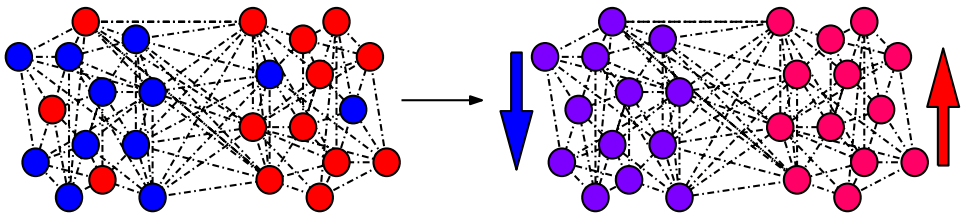

FIGURE 12. The typical behavior of the Averaging dynamics on a graph which exhibits a "good" community structure, i.e. where the second eigenvector is close to the characteristic vector of the two communities and the third eigenvalue is smaller than the second one by a constant factor. The internal expansion of the two communities leads the values of pairs of nodes in a community to be much closer than those of pairs of nodes in different communities. Once such configuration is reached, the edges in the cut make the nodes slowly converge to a common value which lies between the averages of the two communities, causing the value of single nodes to evolve monotonically.

Let $G = (V, E)$ be an undirected graph (possibly with multiple edges and self loops), A its adjacency matrix and d_i the degree of node i. The

[2]We adopt the common convention that $f(S) := \{f(x) : x \in S\}$ for any function f with domain D and any subset $S \subseteq D$.

transition matrix of (the random walk on) G is the matrix

$$P = D^{-1}A,$$

where D is the diagonal matrix such that $D_{i,i} = d_i$. $P_{i,j} = (1/d_i) \cdot A_{i,j}$ is thus the probability of going from i to j in one-step of the random walk on G. P operates as the random walk process on G by left multiplication, and as the Averaging dynamics by right multiplication. For $i = 1, 2$, define $\mathbf{1}_{V_i}$, as the $|V|$-dimensional vector, whose j-th component is 1 if $j \in V_i$, it is 0 otherwise. If (V_1, V_2) is a bipartition of the nodes with $|V_1| = |V_2| = n$, we define the *partition indicator vector*

$$\boldsymbol{\chi} = \mathbf{1}_{V_1} - \mathbf{1}_{V_2}.$$

If $\mathbf{x}$ is the initial vector of values, after t rounds of the Averaging dynamics the vector of values at time t is

$$\mathbf{x}^{(t)} = P^t \mathbf{x}.$$

The product of the power of a matrix times a vector is best understood in terms of the spectrum of the matrix, which is what we explore in the next section.

In what follows we always denote by $\lambda_1 \geqslant \ldots \geqslant \lambda_{2n}$ the eigenvalues of P. Recall that, since P is a stochastic matrix we have $\lambda_1 = 1$ and $\lambda_{2n} \geqslant -1$, moreover for all graphs that are connected and not bipartite it holds that $\lambda_2 < 1$ and $\lambda_{2n} > -1$. We denote by λ the largest, in absolute value, among all but the first two eigenvalues, namely

$$\lambda = \max\{|\lambda_i| \,:\, i = 3, 4, \ldots, 2n\}.$$

4.3. Length of the Projection of Vector x

For the analysis of the Averaging dynamics on both regular and non-regular graphs, it is important to understand the distribution of the projection of $\mathbf{x}$ on $\mathbf{1}$ and $\boldsymbol{\chi}$, that is (up to scaling) the distribution of the inner products $\langle \mathbf{x}, \mathbf{1} \rangle$ and $\langle \mathbf{x}, \boldsymbol{\chi} \rangle$. In particular we are going to use the following bound.

LEMMA 2. *If we pick* $\mathbf{x}$ *uniformly at random in* $\{-1, 1\}^{2n}$ *then, for any* $\delta > 0$ *and any fixed vector* $\mathbf{w} \in \{-1, 1\}^{2n}$ *with* ± 1 *entries, it holds*

$$\Pr\left(\left|\langle (1/\sqrt{2n})\,\mathbf{w},\, \mathbf{x}\rangle\right| \leqslant \delta\right) \leqslant \mathcal{O}(\delta).$$

PROOF. Since $\mathbf{x}$ is a vector of independent and uniformly distributed random variables in $\{-1, 1\}$, both $\langle \mathbf{x}, \boldsymbol{\chi} \rangle$ and $\langle \mathbf{x}, \mathbf{1} \rangle$ have the distribution of a sum of $2n$ Rademacher random variables[3]. Such a sum takes the value $2k - 2n$ with probability $\frac{1}{2^n}\binom{2n}{k}$, and so every possible value has probability

[3]A Rademacher random variable X is such that $\Pr(X = +1) = \Pr(X = -1) = \frac{1}{2}$.

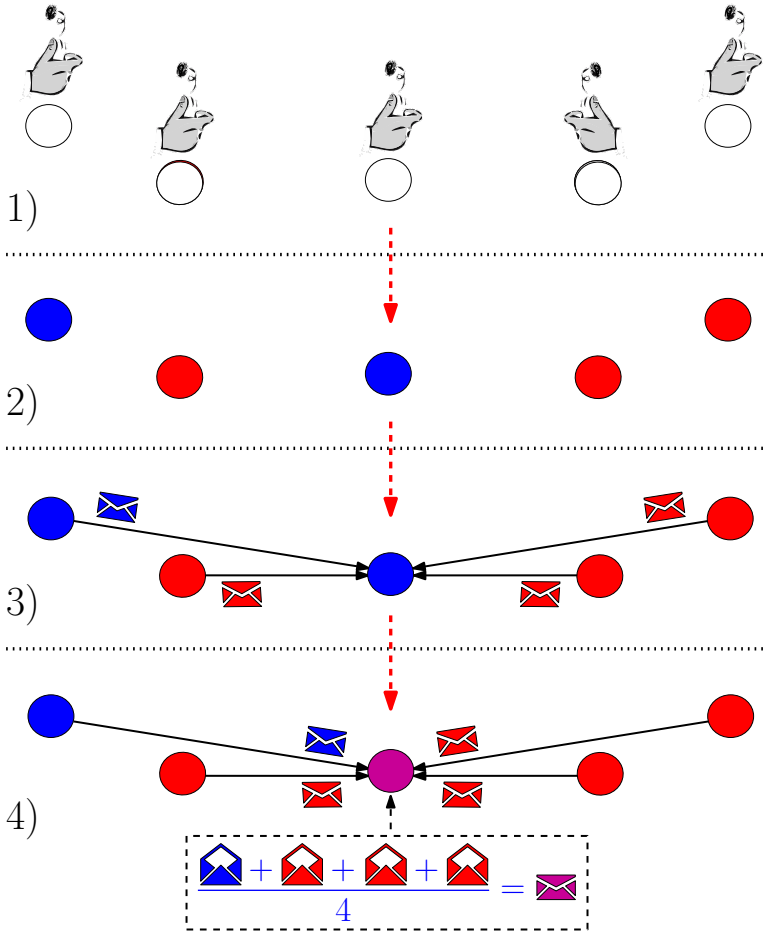

FIGURE 13. A pictorial representation of the Rademacher initialization and the application of the Averaging dynamics (step (1) of the updating rule in Algorithm 1): **1)-2):** The nodes generate a random variable in $\{-1,+1\}$ u.a.r. **3)-4):** Each node sends her current value to all the neighbors, and updates her value with the average of those received from the neighbors.

at most $\frac{1}{2^n}\binom{2n}{n} \approx \frac{1}{\sqrt{2\pi n}}$. Consequently, if R is the sum of $2n$ Rademacher random variables, we have

$$\Pr\left(|R| \leqslant \delta\sqrt{2n}\right) \leqslant \mathcal{O}(\delta).$$

gg

Although it is possible to argue that a Rademacher vector has $\Omega(1)$ probability of having inner product $\Omega(\|\mathbf{w}\|)$ with every vector $\mathbf{w}$, such a statement does not hold w.h.p. We do have, however, estimates of the inner

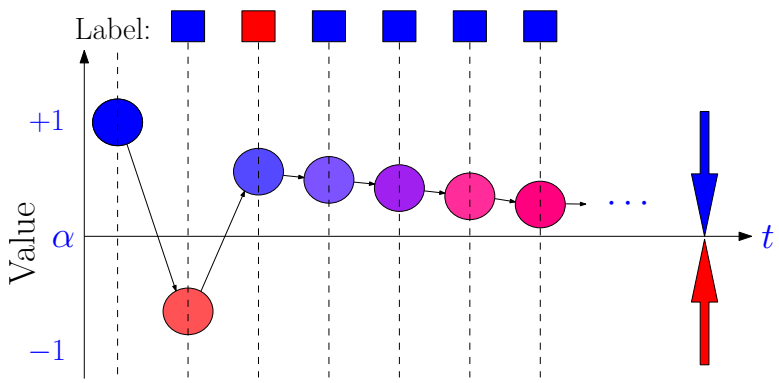

FIGURE 14. A pictorial representation of the labeling criterion of the Averaging protocol (step (2) of the updating rule in Algorithm 1): nodes whose value increases from one round to the next label themselves "red", otherwise they label themselves "blue".

product of a vector $\mathbf{w}$ with a Rademacher vector $\mathbf{x}$ provided that $\mathbf{w}$ is close to a vector in $\{-1,1\}^{2n}$.

LEMMA 3. *Let k be a positive integer. For every nk-dimensional vector $\mathbf{w}$ such that*

$$|\{i \mid |\mathbf{w}(i)| \geqslant c\}| \geqslant n,$$

for some positive constant c, if we pick $\mathbf{x}$ uniformly at random in $\{-1,1\}^{kn}$, then

$$\Pr\left(\left|\langle \frac{1}{\sqrt{kn}}\mathbf{w}, \mathbf{x}\rangle\right| \leqslant \delta\right) \leqslant \mathcal{O}(k\delta) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

PROOF. Let $S \subset \{1,\ldots,kn\}$ be the set of coordinates i of $\mathbf{w}$ such that $|\mathbf{w}(i)| \geqslant c$. By hypothesis, we have $|S| \geqslant n$. Let $T := \{1,\ldots,kn\} - S$.

Next, for every assignment $\mathbf{a} \in \{-1,1\}^{kn}$, we show that

$$\Pr\left(|\langle \mathbf{w}, \mathbf{x}\rangle| \leqslant \delta\sqrt{kn} \mid \ \forall i \in T,\, \mathbf{x}(i) = \mathbf{a}(i)\right) \leqslant \mathcal{O}(\delta),$$

from which the lemma follows.

Call $t := \sum_{i\in T} a_i z_i$. We need to show

$$\Pr\left(|\sum_{i\in S} \mathbf{x}(i)\mathbf{w}(i) + t| \leqslant \delta\sqrt{kn}\right) \leqslant \mathcal{O}(\delta).$$

From the Berry-Esseen theorem,

$$\Pr\left(|\sum_{i\in S} \mathbf{x}(i)\mathbf{w}(i) + t| \leqslant \delta\sqrt{kn}\right) \leqslant \Pr\left(|g + t| \leqslant \delta\sqrt{kn}\right) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right),$$

where g is a Gaussian random variable of mean 0 and variance

$$\sigma^2 = \sum_{i \in S} (\mathbf{w}(i))^2 \geqslant c^2 |S| \geqslant c^2\, n,$$

thus

$$\Pr\left(|g+t| \leqslant \delta\sqrt{kn}\right) = \frac{1}{\sqrt{2\sigma^2\pi}} \int_{-t-\delta\sqrt{kn}}^{-t+\delta\sqrt{kn}} e^{-\frac{s^2}{2\sigma^2}}\, ds \leqslant \frac{2\delta\sqrt{kn}}{\sqrt{2\pi c^2\, n}} = \frac{\sqrt{2k}\delta}{\sqrt{\pi}c},$$

where we used the fact that $e^{-s^2/2} \leqslant 1$ for all s, concluding the proof. gs

4.4. Strong Reconstruction for Regular Graphs

Observe that, if G is d-regular then $P = (1/d) \cdot A$ is a real symmetric matrix and P and A have the same set of eigenvectors. We denote by $\mathbf{v}_1 = (1/\sqrt{2n})\mathbf{1}, \mathbf{v}_2, \ldots, \mathbf{v}_{2n}$ a basis of orthonormal eigenvectors, where each $\mathbf{v}_i$ is the eigenvector associated to eigenvalue λ_i. Then, we can write a vector $\mathbf{x}$ as a linear combination $\mathbf{x} = \sum_i \alpha_i \mathbf{v}_i$ and we have:

$$P^t\mathbf{x} = \sum_i \lambda_i^t \alpha_i \mathbf{v}_i = \frac{1}{2n}\left(\sum_i \mathbf{x}(i)\right)\mathbf{1} + \sum_{i=2}^{2n} \lambda_i^t \alpha_i \mathbf{v}_i,$$

which implies that $\mathbf{x}^{(t)} = P^t\mathbf{x}$ tends to $\alpha_1\mathbf{v}_1$ as t tends to infinity, i.e., it converges to the vector that has the average of $\mathbf{x}$ in every coordinate.

We next show that, if the regular graph is "well" clustered, then the Averaging protocol produces a strong reconstruction of the two clusters, w.h.p.

Definition 3 (Clustered Regular Graph). A $(2n, d, b)$-clustered regular graph $G = ((V_1, V_2), E)$ is a connected graph over node set $V_1 \cup V_2$, with $|V_1| = |V_2| = n$ and such that:

- Every node has degree d,
- Every node in cluster V_1 has b neighbors in cluster V_2 and every node in V_2 has b neighbors in V_1.

Let $G = ((V_1, V_2), E)$ be a $(2n, d, b)$-clustered regular graph with adjacency matrix A and transition matrix $P = (1/d) \cdot A$.

We know that $\mathbf{1}$ is an eigenvector of P with eigenvalue 1. Furthermore, the partition indicator vector $\boldsymbol{\chi}$ is an eigenvector of P with eigenvalue $1 - 2b/d$, as given by the following observation.

Observation 2. If G is a $(2n, d, b)$-clustered regular graph with clusters V_1 and V_2 and $\boldsymbol{\chi} = \mathbf{1}_{V_1} - \mathbf{1}_{V_2}$ is the partition indicator vector, then $\boldsymbol{\chi}$ is an eigenvector of the transition matrix P of G with eigenvalue $1 - 2b/d$.

Proof. Every node i has b neighbors j on the opposite side of the partition, for which $\boldsymbol{\chi}(j) = -\boldsymbol{\chi}(i)$, and $d - b$ neighbors j on the same side, for which $\boldsymbol{\chi}(j) = \boldsymbol{\chi}(i)$, so

$$(P\boldsymbol{\chi})_i = \frac{1}{d}\left((d-b)\boldsymbol{\chi}(i) - b\boldsymbol{\chi}(i)\right) = \left(1 - \frac{2b}{d}\right)\boldsymbol{\chi}(i).$$

We first show that, if $1-2b/d$ happens to be the second eigenvalue, after t rounds of the Averaging dynamics, the configuration $\mathbf{x}^{(t)}$ is close to a linear combination of $\mathbf{1}$ and $\boldsymbol{\chi}$. Formally, if $\lambda < 1-2b/d$ the following holds.

LEMMA 4. *Assume we run the Averaging dynamics in a $(2n,d,b)$-clustered regular graph G (see Definition 3) with any initial vector $\mathbf{x} \in \{-1,1\}^{2n}$. If $\lambda < 1-2b/d$ then there are reals α_1, α_2 such that at every round t we have*

$$\mathbf{x}^{(t)} = \alpha_1 \mathbf{1} + \alpha_2 \lambda_2^t \boldsymbol{\chi} + \mathbf{e}^{(t)} \quad \text{where} \quad \left\|\mathbf{e}^{(t)}\right\|_\infty \leqslant \lambda^t \sqrt{2n}\,. \tag{1}$$

PROOF. Since $\mathbf{x}^{(t)} = P^t \mathbf{x}$ we can write

$$P^t \mathbf{x} = \sum_i \lambda_i^t \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i,$$

where $1 = \lambda_1 > \lambda_2 = 1-2b/d > \lambda_3 \geqslant \cdots \geqslant \lambda_{2n}$ are the eigenvalues of P and $\mathbf{v}_1 = \frac{1}{\sqrt{2n}}\mathbf{1}$, $\mathbf{v}_2 = \frac{1}{\sqrt{2n}}\boldsymbol{\chi}$, $\mathbf{v}_3$, $\ldots$, $\mathbf{v}_{2n}$ are a corresponding sequence of orthonormal eigenvectors. Hence,

$$\begin{aligned}\mathbf{x}^{(t)} &= \frac{1}{2n}\langle \mathbf{x}, \mathbf{1}\rangle \cdot \mathbf{1} + \lambda_2^t \frac{1}{2n}\langle \mathbf{x}, \boldsymbol{\chi}\rangle \cdot \boldsymbol{\chi} + \sum_{i=3}^{2n} \lambda_i^t \alpha_i \mathbf{v}_i \\ &= \alpha_1 \mathbf{1} + \alpha_2 \lambda_2^t \cdot \boldsymbol{\chi} + \sum_{i=3}^{2n} \lambda_i^t \alpha_i \mathbf{v}_i,\end{aligned}$$

where we set $\alpha_1 = \frac{1}{2n}\langle \mathbf{1}, \mathbf{x}\rangle$ and $\alpha_2 = \frac{1}{2n}\langle \boldsymbol{\chi}, \mathbf{x}\rangle$. We bound the ℓ_∞ norm of the last term as

$$\begin{aligned}\left\|\sum_{i=3}^{2n} \lambda_i^t \alpha_i \mathbf{v}_i\right\|_\infty &\leqslant \left\|\sum_{i=3}^{2n} \lambda_i^t \alpha_i \mathbf{v}_i\right\|_2 = \sqrt{\sum_{i=3}^{2n} \lambda_i^{2t} \alpha_i^2} \\ &\leqslant \lambda^t \sqrt{\sum_{i=1}^{2n} \alpha_i^2} = \lambda^t \|\mathbf{x}\| = \lambda^t \sqrt{2n}.\end{aligned}$$

Informally speaking, (1) naturally "suggested" the choice of the coloring rule in the Averaging protocol, once we considered the difference of two consecutive values of any node u, i.e.,

$$\mathbf{x}^{(t-1)}(u) - \mathbf{x}^{(t)}(u) = \alpha_2 \lambda_2^{t-1}(1-\lambda_2)\boldsymbol{\chi}(u) + \mathbf{e}^{(t-1)}(u) - \mathbf{e}^{(t)}(u)\,. \tag{2}$$

(See Figure 15 for an interpretation of α_1, α_2.) Intuitively, if λ is sufficiently small, we can exploit the bound on $\left\|\mathbf{e}^{(t)}\right\|_\infty$ in (1) to show that, after a short initial phase, the sign of $\mathbf{x}^{(t-1)}(u) - \mathbf{x}^{(t)}(u)$ is essentially determined by $\boldsymbol{\chi}(u)$, thus by the community u belongs to, w.h.p. The following theorem and its proof formalize the above fact.

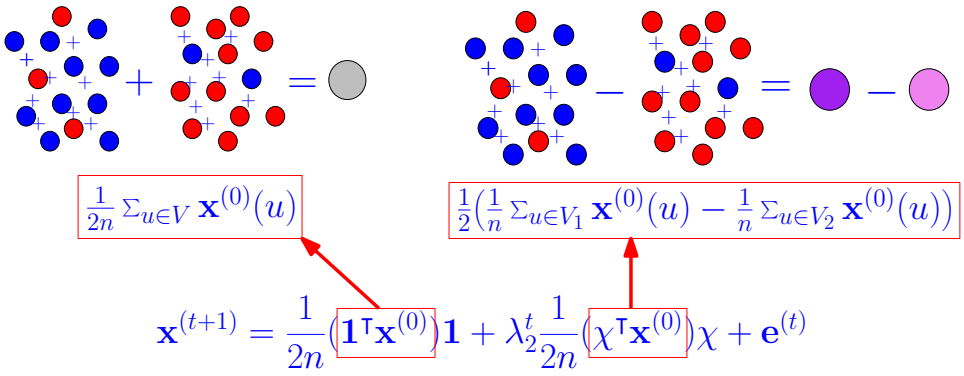

FIGURE 15. An illustration of the interpretation of the projections on the first and second eigenvectors of the adjacency matrix in the regular case: the first projection is the *global* average of the initial values in the whole graph, while the second one is the difference between the averages of the initial values *within the two communities*.

THEOREM 1 (Strong Reconstruction). *Let $G=((V_1,V_2),E)$ be a connected $(2n,d,b)$-clustered regular graph with $1-2b/d>(1+\delta)\lambda$ for an arbitrarily-small constant $\delta>0$. Then the Averaging protocol produces a strong reconstruction within $\mathcal{O}(\log n)$ rounds, w.h.p.*

SKETCH OF PROOF. From (2), we have that

$$\mathtt{sgn}\left(\mathbf{x}^{(t-1)}(u)-\mathbf{x}^{(t)}(u)\right)=\mathtt{sgn}\left(\alpha_2\boldsymbol{\chi}(u)\right)$$

whenever

$$\left|\alpha_2\lambda_2^{t-1}(1-\lambda_2)\right|>\left|\mathbf{e}^{(t-1)}(u)-\mathbf{e}^{(t)}(u)\right|. \tag{3}$$

From (1) we have that

$$\left|\mathbf{e}^{(t)}(u)\right|\leqslant\lambda^t\sqrt{2n},$$

thus (3) is satisfied for all t such that

$$t-1\geqslant\log\left(\frac{2\sqrt{2n}}{|\alpha_2|(1-\lambda_2)}\right)\cdot\frac{1}{\log\left(\lambda_2/\lambda\right)}.$$

The second key-step of the proof relies on the randomness of the initial vector. Indeed, since $\mathbf{x}$ is a vector of independent and uniformly distributed random variables in $\{-1,1\}$, the absolute difference between the two partial averages in the two communities, i.e. $|\alpha_2|$, is "sufficiently" large, w.h.p. More precisely, from Lemma 2 we have that is the sum of $2n$ Rademacher random variables, we have

$$\Pr\left(|R|\leqslant\delta\sqrt{2n}\right)\leqslant\mathcal{O}(\delta).$$

Since $\alpha_2 = \frac{1}{2n}\langle \boldsymbol{\chi}, \mathbf{x}\rangle$ and $\mathbf{x}$ is a vector of Rademacher random variables, the previous inequality implies that

$$|\alpha_2| = \frac{1}{2n}\langle \boldsymbol{\chi}, \mathbf{x}\rangle \geqslant n^{-\gamma},$$

for some positive constant γ, w.h.p. The theorem thus follows from the above bound on $|\alpha_2|$ and from the hypothesis $\lambda_2 \geqslant (1+\delta)\lambda$. ▨

REMARK 3. Graphs to which Theorem 1 apply are those consisting of two regular expanders connected by a regular sparse cut. Indeed, let $G = ((V_1, V_2), E)$ be a $(2n, d, b)$-clustered regular graph, and let $\lambda_A = \max\{\lambda_2(A_1), \lambda_2(A_2)\}$ and $\lambda_B = \lambda_2(B)$, where A_1, A_2 and B are the adjacency matrices of the subgraphs induced by V_1, V_2 and the cut between V_1 and V_2, respectively. Since

$$\lambda = \frac{a}{d}\lambda_A + \frac{b}{d}\lambda_B,$$

if

$$a - b > (1+\varepsilon)(a\lambda_A + b\lambda_B),$$

G satisfies the hypothesis of Theorem 1.

4.4.1. Regular stochastic block model

We can use Theorem 1 to prove that the Averaging protocol achieves strong reconstruction in the regular stochastic block model [**BDG$^+$16**], defined as follows.

DEFINITION 4 (Regular Stochastic Block Model). In the regular stochastic block model with two communities, a graph on $2n$ nodes is obtained as follows: Given two parameters $a(n)$ and $b(n)$ (*internal* and *external* degrees, respectively), partition nodes into two equal-sized subsets V_1 and V_2 and then sample a random $a(n)$-regular graph over each of V_1 and V_2 and a random $b(n)$-regular graph between V_1 and V_2.

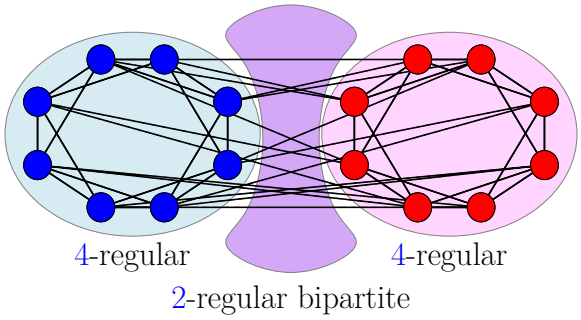

FIGURE 16. An example of a regular stochastic block model (Definition 4) with $n = 8$, $a = 4$ and $b = 2$.

If G is a graph sampled from the regular stochastic block model with internal and external degrees a and b respectively, then it is a $(2n, d, b)$-clustered graph with largest eigenvalue of the transition matrix 1 and corresponding eigenvector $\mathbf{1}$, while $\boldsymbol{\chi}$ is also an eigenvector, with eigenvalue $1 - 2b/d$, where $d := a + b$. Furthermore, we can derive the following upper bound on the maximal absolute value achieved by the other $2n-2$ eigenvalues corresponding to eigenvectors orthogonal to $\mathbf{1}$ and $\boldsymbol{\chi}$:

$$\lambda \leqslant \frac{2}{a+b}(\sqrt{a+b-1} + o_n(1)) \tag{4}$$

This bound can be proved using some general result of Friedman and Kohler [**FK14**] on *random degree k lifts* of a graph, as given in the following.

LEMMA 5. *Let G be a graph sampled from the regular stochastic block model with internal and external degrees a and b respectively. It holds that w.h.p.*

$$\lambda \leqslant \frac{2}{a+b}(\sqrt{a+b-1} + o_n(1)).$$

SKETCH OF PROOF. The lemma follows from the general results of Friedman and Kohler [**FK14**], simplified by Bordenave [**Bor15b**]. If G is a multigraph on n nodes, then a *random degree k lift* of G is a distribution over graphs G' on kn nodes sampled as follows: every node v of G is replaced by k nodes $v_1, \ldots, v_k$ in G', every edge (u, v) in G is replaced by a random bipartite matching between $u_1, \ldots, u_k$ and $v_1, \ldots, v_k$ (if there are multiple edges, each edge is replaced by an independently sampled matching) and every self loop over u is replaced by a random degree-2 graph over $u_1, \ldots, u_k$ which is sampled by taking a random permutation

$$\pi : \{1, \ldots, k\} \to \{1, \ldots, k\}$$

and connecting u_i to $u_\pi(i)$ for every i.

For every lift of any d-regular graph, the lifted graph is still d-regular, and every eigenvalue of the adjacency matrix of the base graph is still an eigenvalue of the lifted graph. Friedman and Kohler [**FK14**] prove that, if $d \geqslant 3$, then with probability $1 - \mathcal{O}(1/k)$ over the choice of a random lift of degree k, the new eigenvalues of the adjacency matrix of the lifted graph are at most $2\sqrt{d-1} + o_k(1)$ in absolute value. Bordenave [**Bor15b**, Corollary 20] has considerably simplified the proof of Friedman and Kohler; although he does not explicitly state the probability of the above event, his argument also bound the failure probability by $1/k^{\Omega(1)}$ [**Bor15a**].

The lemma now follows by observing that the regular stochastic block model is a random lift of degree n of the graph that has only two nodes v_1 and v_2, it has b parallel edges between v_1 and v_2, and it has $a/2$ self-loops on v_1 and $a/2$ self-loops on v_2. ∎

From Lemma 5, since $\lambda_2 = \frac{a-b}{a+b}$, using (4) in Theorem 1, we get a strong reconstruction for the regular stochastic block model.

COROLLARY 1 (Reconstruction in Regular Stochastic Block Models). *Let G be a random graph sampled from the regular stochastic block model with*

$$a - b > 2(1+\eta)\sqrt{a+b}$$

for an arbitrarily small constant $\eta > 0$, then the Averaging protocol produces a strong reconstruction in $\mathcal{O}(\log n)$ rounds, w.h.p.

4.5. Weak Reconstruction for Non-Regular Graphs

The results of Section 4.4 rely on very clear spectral properties of regular, clustered graphs, immediately reflecting their underlying topological structure. Intuition suggests that these properties should be approximately preserved if we suitably relax the notion of regularity. We thus generalize our approach to a large class of non-regular clustered graphs.

DEFINITION 5 (Clustered γ-Regular Graphs). A $(2n, d, b, \gamma)$-clustered graph $G = ((V_1, V_2), E)$ (with $\gamma < 1$), is a graph over node set $V_1 \cup V_2$, where $|V_1| = |V_2| = n$ such that:

- Every node has degree $d \pm \gamma d$,
- Every node in V_1 has $b \pm \gamma d$ neighbors in V_2 and every node in V_2 has $b \pm \gamma d$ neighbors in V_1.

If G is not regular then the matrix $P = D^{-1}A$ is not symmetric in general, however it is possible to relate its eigenvalues and eigenvectors to those of a symmetric matrix, as follows. Denote the *normalized adjacency matrix* of G as

$$N := D^{-1/2}AD^{-1/2} = D^{1/2}PD^{-1/2}.$$

Notice that N is symmetric, P and N have the same eigenvalues $\lambda_1, \ldots, \lambda_{2n}$, and $\mathbf{x}$ is an eigenvector of P if and only if $D^{1/2}\mathbf{x}$ is an eigenvector of N (if G is regular then P and N are the same matrix). Let $\mathbf{w}_1, \ldots, \mathbf{w}_{2n}$ be a basis of orthonormal eigenvectors of N, with $\mathbf{w}_i$ the eigenvector associated to eigenvalue λ_i, for every i. We have that

$$\mathbf{w}_1 = \frac{D^{1/2}\mathbf{1}}{\|D^{1/2}\mathbf{1}\|}.$$

If we set $\mathbf{v}_i := D^{-1/2}\mathbf{w}_i$, we obtain a set of eigenvectors for P and we can write $\mathbf{x} = \sum_i \alpha_i \mathbf{v}_i$ as a linear combination of them. Then, the averaging process can again be described as

$$P^t\mathbf{x} = \sum_i \lambda_i^t \alpha_i \mathbf{v}_i = \alpha_1\mathbf{v}_1 + \sum_i \lambda_i^t \alpha_i \mathbf{v}_i.$$

So, if G is connected and not bipartite, the Averaging dynamics converges to $\alpha_1\mathbf{v}_1$. In general, it is easy to see that $\alpha_i = \mathbf{w}_i^T D^{1/2}\mathbf{x}$ (see the first lines in the proof of Lemma 6) and $\alpha_1\mathbf{v}_1$ is the vector

$$(\mathbf{w}_1^T D^{1/2}\mathbf{x}) \cdot D^{-1/2}\mathbf{w}_1 = \frac{\mathbf{1}^T D\mathbf{x}}{\|D^{1/2}\mathbf{1}\|^2}\mathbf{1} = \frac{\sum_i d_i \mathbf{x}(i)}{\sum_i d_i} \cdot \mathbf{1}\,.$$

As in the regular case, if the transition matrix P of a clustered γ-regular graph has λ_2 close to 1 and $|\lambda_3|, \ldots, |\lambda_{2n}|$ small, the Averaging dynamics has a long phase in which $\mathbf{x}^{(t)} = P^t\mathbf{x}$ is close to $\alpha_1\mathbf{1} + \alpha_2\mathbf{v}_2$. However, providing an argument similar to the regular case is considerably harder, since the partition indicator vector $\boldsymbol{\chi}$ is no longer an eigenvector of P. In order to fix this issue, we generalize (1), proving in Lemma 6 that $\mathbf{x}^{(t)}$ is still close to a linear combination of $\mathbf{1}$ and $\boldsymbol{\chi}$. We set $\nu = 1 - \frac{2b}{d}$, since this value occurs frequently in this section.

LEMMA 6. *Let Averaging dynamics run on a a connected $(2n, d, b, \gamma)$-clustered graph G with $\gamma \leqslant 1/10$, with initial vector $\mathbf{x}$. If $\lambda < \nu$ we have:*

$$\mathbf{x}^{(t)} = \alpha_1\mathbf{1} + \alpha_2\lambda_2^t\boldsymbol{\chi} + \alpha_2\lambda_2^t\mathbf{z} + \mathbf{e}^{(t)},$$

for some vectors $\mathbf{z}$ and $\mathbf{e}^{(t)}$ with

$$\|\mathbf{z}\| \leqslant \frac{88\,\gamma}{\nu - \lambda_3}\sqrt{2n} \quad \text{and} \quad \|\mathbf{e}^{(t)}\| \leqslant 4\lambda^t\|\mathbf{x}\|.$$

Coefficients α_1 and α_2 are

$$\|\mathbf{z}\| \leqslant \frac{88\,\gamma}{\nu - \lambda_3}\sqrt{2n} \quad \text{and} \quad \|\mathbf{e}^{(t)}\| \leqslant 4\lambda^t\|\mathbf{x}\|.$$

PROOF. We prove the following two key-facts:

(i) the second eigenvalue of the transition matrix of G is not much smaller than $1 - 2b/d$, and
(ii) $D^{1/2}\boldsymbol{\chi}$ is close, in norm, to its projection on the second eigenvector of the normalized adjacency matrix N.

Namely, in Lemma 10 we prove that if $\lambda_3 < \nu$ then

$$\lambda_2 \geqslant \nu - 10\gamma \quad \text{and} \quad \left\|D^{1/2}\boldsymbol{\chi} - \beta_2\mathbf{w}_2\right\| \leqslant \frac{44\,\gamma}{\nu - \lambda_3}\sqrt{2nd}, \tag{5}$$

where $\beta_2 = \boldsymbol{\chi}^\intercal D^{1/2}\mathbf{w}_2$.

Now, we can use the above bounds to analyze $\mathbf{x}^{(t)} = P^t\mathbf{x}$. To begin, note that

$$N = D^{-1/2}AD^{-1/2} \quad \text{and} \quad P = D^{-1}A$$

imply that

$$P = D^{-1/2}ND^{1/2} \quad \text{and} \quad P^t = D^{-1/2}N^tD^{1/2}.$$

Thus, for any vector $\mathbf{x}$, if we write $D^{1/2}\mathbf{x}$ as a linear combination of an orthonormal basis of N,

$$D^{1/2}\mathbf{x} = \sum_{i=1}^{2n} a_i\mathbf{w}_i,$$

we get

$$P^t\mathbf{x} = D^{-1/2}N^tD^{1/2}\mathbf{x} = D^{-1/2}\sum_{i=1}^{2n} a_i\lambda_i^t\mathbf{w}_i = \sum_{i=1}^{2n} a_i\lambda_i^tD^{-1/2}\mathbf{w}_i. \tag{6}$$

We next estimate the first term, the second term, and the sum of the remaining terms of (6).

First term of (6)**.** We have $\mathbf{w}_1 = \frac{D^{1/2}\mathbf{1}}{\|D^{1/2}\mathbf{1}\|}$, so the first term can be written as $\alpha_1\mathbf{1}$ with

$$\alpha_1 = \frac{a_1}{\left\|D^{1/2}\mathbf{1}\right\|} = \frac{\mathbf{w}_1^\intercal D^{1/2}\mathbf{x}}{\left\|D^{1/2}\mathbf{1}\right\|} = \frac{\mathbf{1}^\intercal D\mathbf{x}}{\left\|D^{1/2}\mathbf{1}\right\|^2}.$$

Second term of (6)**.** If we write

$$D^{1/2}\boldsymbol{\chi} = \beta_2\mathbf{w}_2 + \mathbf{y},$$

with $\beta_2 = \mathbf{w}_2^\intercal D^{1/2}\boldsymbol{\chi}$, (5) implies that

$$\|\mathbf{y}\| \leqslant \frac{44\,\gamma}{\nu - \lambda_3}\sqrt{2nd}.$$

Hence the second term can be written as

$$\begin{aligned} a_2\lambda_2^t D^{-1/2}\mathbf{w}_2 &= a_2\lambda_2^t D^{-1/2}\left(\frac{D^{1/2}\boldsymbol{\chi} - \mathbf{y}}{\beta_2}\right) \\ &= \frac{a_2}{\beta_2}\lambda_2^t\boldsymbol{\chi} - \frac{a_2}{\beta_2}\lambda_2^t\mathbf{z} = \alpha_2\lambda_2^t\boldsymbol{\chi} - \alpha_2\lambda_2^t\mathbf{z}, \end{aligned}$$

where

$$\|\mathbf{z}\| = \left\|D^{-1/2}\mathbf{y}\right\| \leqslant \left\|D^{-1/2}\right\|\,\|\mathbf{y}\| \leqslant \frac{2}{\sqrt{d}}\cdot\frac{44\,\gamma}{\nu-\lambda_3}\sqrt{2nd} = \frac{88\,\gamma}{\nu-\lambda_3}\sqrt{2n},$$

and

$$\alpha_2 = a_2/\beta_2 = \frac{\mathbf{w}_2^\intercal D^{1/2}\mathbf{x}}{\mathbf{w}_2 D^{1/2}\boldsymbol{\chi}}\,.$$

Remaining terms of (6)**.** As for all other terms, observe that

$$\begin{aligned} \|\mathbf{e}^{(t)}\|^2 &= \left\|\sum_{i=3}^{2n} a_i\lambda_i^t D^{-1/2}\mathbf{w}_i\right\|^2 \\ &\leqslant \left\|D^{-1/2}\right\|^2 \left\|\sum_{i=3}^{2n} a_i\lambda_i^t\mathbf{w}_i\right\|^2 \\ &= \left\|D^{-1/2}\right\|^2 \sum_{i=3}^{2n} a_i^2\lambda_i^{2t} \\ &\leqslant \left\|D^{-1/2}\right\|^2 \lambda^{2t}\sum_{i=3}^{2n} a_i^2 \\ &\leqslant \left\|D^{-1/2}\right\|^2 \lambda^{2t}\left\|D^{1/2}\mathbf{x}\right\|^2 \\ &\leqslant \left\|D^{-1/2}\right\|^2 \left\|D^{1/2}\right\|^2 \lambda^{2t}\|\mathbf{x}\|^2 \leqslant 16\lambda^{2t}\|\mathbf{x}\|^2. \end{aligned}$$

gg

The above lemma allows us to generalize our approach to achieve efficient, weak reconstruction in non-regular clustered graphs.

THEOREM 3 (Weak Reconstruction). *Let G be a connected $(2n, d, b, \gamma)$-clustered graph with $\gamma \leqslant c(\nu - \lambda_3)$ for a suitable constant $c > 0$. If $\lambda < \nu$ and $\lambda_2 \geqslant (1+\delta)\lambda$ for an arbitrarily-small positive constant δ, then the Averaging protocol produces an $\mathcal{O}(\gamma^2/(\nu - \lambda_3)^2)$-weak reconstruction within $\mathcal{O}(\log n)$ rounds, w.h.p.*[4]

PROOF. Lemma 6 implies that for every node u at any round t we have

$$\mathbf{x}^{(t-1)}(u) - \mathbf{x}^{(t)}(u) = \alpha_2 \lambda_2^{t-1}(1-\lambda_2)\left(\boldsymbol{\chi}(u) + \mathbf{z}(u)\right) + \mathbf{e}^{(t-1)}(u) - \mathbf{e}^{(t)}(u).$$

Hence, for every node u such that $|\mathbf{z}(u)| < 1/2$,[5] we have

$$\mathtt{sgn}\left(\mathbf{x}^{(t-1)}(u) - \mathbf{x}^{(t)}(u)\right) = \mathtt{sgn}\left(\alpha_2 \boldsymbol{\chi}(u)\right)$$

whenever

$$\left|\frac{1}{2}\alpha_2 \lambda_2^{t-1}(1-\lambda_2)\right| > \left|\mathbf{e}^{(t-1)}(u) - \mathbf{e}^{(t)}(u)\right|. \tag{7}$$

From Lemma 6 we have $\left|\mathbf{e}^{(t)}(u)\right| \leqslant 4\lambda^t\sqrt{2n}$, thus (7) is satisfied for any t such that

$$t - 1 \geqslant \log\left(\frac{16\sqrt{2n}}{|\alpha_2|(1-\lambda_2)}\right) \cdot \frac{1}{\log\left(\lambda_2/\lambda\right)}. \tag{8}$$

The right-hand side in the above formula is $\mathcal{O}(\log n)$, w.h.p., because of the following three points:

- From Cheeger's inequality (see e.g. [**Chu96**]) and the fact that the graph is connected it follows that $1 - \lambda_2 \geqslant 1/(2n^4)$;
- $\lambda_2 \geqslant (1+\delta)\lambda$ by hypothesis;
- It holds w.h.p. $|\alpha_2| \geqslant n^{-c}$ for some large enough positive constant c, as a consequence of the following equations that we prove below:

$$\begin{aligned}\Pr\left(|\alpha_2| \leqslant \frac{1}{n^c}\right) &= \Pr\left(\frac{\left|\mathbf{w}_2^\intercal D^{\frac{1}{2}}\mathbf{x}\right|}{\left|\mathbf{w}_2^\intercal D^{\frac{1}{2}}\boldsymbol{\chi}\right|} \leqslant \frac{1}{n^c}\right) \\ &\leqslant \Pr\left(\left|\mathbf{w}_2^\intercal D^{1/2}\mathbf{x}\right| \leqslant \frac{2\sqrt{d}}{n^{c-1/2}}\right) \leqslant \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).\end{aligned} \tag{9}$$

In the first inequality of (9) we used that, by definition,

$$|\alpha_2| = |\mathbf{w}_2^\intercal D^{\frac{1}{2}}\mathbf{x}| / |\mathbf{w}_2^\intercal D^{\frac{1}{2}}\boldsymbol{\chi}|.$$

[4]Consistently, Theorem 1 is a special case of this one when $\gamma = 0$.

[5]The value $1/2$ is chosen here only for readability sake, any constant smaller than 1 will do.

In the first inequality we used that, by the Cauchy-Schwarz inequality,

$$|\mathbf{w}_2^{\intercal} D^{\frac{1}{2}} \boldsymbol{\chi}| \leqslant \|D^{\frac{1}{2}} \boldsymbol{\chi}\| \leqslant 2\sqrt{dn}.$$

In order to prove the last inequality of (9), we use that from Lemma 10 it holds

$$\begin{aligned}
&\left\| D^{1/2}\boldsymbol{\chi} - \beta_2 \mathbf{w}_2 \right\|^2 \\
&= \left\| D^{1/2}\boldsymbol{\chi} \right\|^2 + \|\beta_2 \mathbf{w}_2\|^2 - 2\langle D^{1/2}\boldsymbol{\chi}, \beta_2 \mathbf{w}_2 \rangle \\
&\leqslant 2\frac{44^2\,\gamma^2}{(\nu-\lambda_3)^2}\, nd,
\end{aligned}$$

that is

$$\begin{aligned}
\langle D^{1/2}\boldsymbol{\chi}, \beta_2 \mathbf{w}_2 \rangle &= \langle D^{1/2}\boldsymbol{\chi}, \mathbf{w}_2 \rangle^2 \\
&\geqslant \frac{1}{2}\left(\left\| D^{1/2}\boldsymbol{\chi} \right\|^2 - 2\frac{44^2\,\gamma^2}{(\nu-\lambda_3)^2}\, nd \right) \geqslant \frac{nd}{3}. \qquad (10)
\end{aligned}$$

Since $\mathbf{w}_2$ is normalized the absolute value of its entries is at most 1, which together with (10) implies that at least a fraction $12/13$ of its entries have an absolute value greater than $1/12$. Thus, we can apply Lemma 3 and prove the last inequality of (9) and, consequently, the fact that (8) is $\mathcal{O}(\log n)$.

Finally, from Lemma 6 we have $\|\mathbf{z}\| \leqslant \frac{88\gamma}{\nu-\lambda_3}\sqrt{2n}$. Thus, the number of nodes u with $\mathbf{z}(u) \geqslant 1/2$ is $\mathcal{O}(n\gamma^2/(\nu-\lambda_3)^2)$. ▣

Roughly speaking, the above theorem states that the quality of the reconstruction depends on the regularity of the graph (through the parameter γ), and the conductance within each community (here represented by the difference $|\nu - \lambda_3|$). Interestingly enough, as long as $|\nu - \lambda_3| = \Theta(1)$, the protocol achieves $\mathcal{O}(\gamma^2)$-weak reconstruction on $(2n, d, b, \gamma)$-clustered graphs.

4.5.1. Reconstruction in the stochastic block model

Below we prove that the stochastic block model $\mathcal{G}_{2n,p,q}$ satisfies the hypotheses of Theorem 3, w.h.p., and, thus, the Averaging protocol efficiently produces a good reconstruction. In what follows, we often use the following parameters of the model: expected internal degree $a = pn$, expected external degree $b = qn$, and $d = a + b$.

LEMMA 7. *Let $G \sim \mathcal{G}_{2n,p,q}$. If $a - b > \sqrt{(a+b)\log n}$ then a positive constant δ exists such that w.h.p.*

i) G is $(2n, d, b, 6\sqrt{\log n/d})$-clustered and
ii) it holds

$$\lambda \leqslant \min\left\{ \frac{\lambda_2}{1+\delta}\,,\; 24\sqrt{\frac{\log n}{d}} \right\}.$$

SKETCH OF PROOF. Claim (i) follows (with probability $1-n^{-1}$) from an easy application of the Chernoff bound (Lemma 76). As for Claim (ii), since G is not regular and random, we derive spectral properties on its adjacency matrix A by considering a "more-tractable" matrix, namely the expected matrix

$$B := \mathbb{E}[A] = \begin{pmatrix} pJ, & qJ \\ qJ, & pJ \end{pmatrix}$$

where $B_{i,j}$ is the probability that the edge (i,j) exists in a random graph $G \sim \mathcal{G}_{2n,p,q}$. In Lemma 8 we prove that such a G is likely to have an adjacency matrix A close to B in spectral norm. Then, in Lemma 9 we show that every clustered graph whose adjacency matrix is close to B has the properties required in the analysis of the Averaging dynamics, thus getting Claim (ii). gg

We now prove Lemma 8 and Lemma 9, which are used in the previous proof of Lemma 7.

LEMMA 8. *If $a(n), b(n)$ are such that $d := a+b > \log n$ and , then w.h.p. (over the choice of $G \sim \mathcal{G}_{2n,\frac{a}{n},\frac{b}{n}}$)*

$$\|A - B\| \leqslant \mathcal{O}(\sqrt{d \log n}).$$

PROOF. We can write $A - B$ as $\sum_{\{i,j\}} X^{\{i,j\}}$, where the matrix $X^{\{i,j\}}$ is zero in all coordinates except (i,j) and (j,i), and, in those coordinates, it is equal to $A - B$. Then we see that the matrices $X^{\{i,j\}}$ are independent, that $\mathbb{E}[X^{\{i,j\}}] = \mathbf{0}$, that $\|X^{\{i,j\}}\| \leqslant 1$ (because every row contains at most one non-zero element, and that element is at most 1 in absolute value), and that $\mathbb{E}[\sum_{\{i,j\}}(X^{\{i,j\}})^2]$ is the matrix that is zero everywhere except for the diagonal entries (i,i) and (j,j), in which we have $B_{i,i} - B_{i,i}^2$ and $B_{j,j} - B_{j,j}^2$ respectively. It follows that

$$\|\mathbb{E}[\sum_{\{i,j\}}(X^{\{i,j\}})^2]\| \leqslant d.$$

Putting these facts together, and applying the Matrix Bernstein Inequality (see Theorem 18 in Section 4.1) with $t = \sqrt{6d \log n}$, we have

$$\Pr\left(\|A - B\| \geqslant \sqrt{9d \log n}\right) \leqslant 2ne^{-\frac{9d \log n}{2d + \frac{2}{3}\sqrt{9d \log n}}} \leqslant 2ne^{-\frac{9d \log n}{4d}} \leqslant 2n^{-1},$$

where we used $d > \log n$. gg

LEMMA 9. *Let G be a $(2n, d, b, \gamma)$-clustered graph such that $\nu = 1 - \frac{2b}{d} > 12\gamma$ and such that its adjacency matrix A satisfies $\|A - B\| \leqslant \gamma d$. Then for every $i \in \{3, \ldots, 2n\}$, $|\lambda_i| \leqslant 4\gamma$ and $\lambda_2 \geqslant (1+\delta)\lambda_3$ for some constant $\delta > 0$.*

PROOF. The matrix B has a very simple spectral structure: $\mathbf{1}$ is an eigenvector of eigenvalue d, $\boldsymbol{\chi}$ is an eigenvector of eigenvalue $a - b$, and all vectors orthogonal to $\mathbf{1}$ and to $\boldsymbol{\chi}$ are eigenvectors of eigenvalue 0. In order to understand the eigenvalues and eigenvectors of N, and hence the eigenvalues

and eigenvectors of P, we first prove that A approximates B and that N approximates $(1/d)A$, namely $\|dN - A\| \leqslant 3\gamma d$.

To show that dN approximates A we need to show that D approximates dI. The condition on the degrees immediately gives us $\|D - dI\| \leqslant \gamma d$. Since every node has degree d_i in the range $d \pm \gamma d$, then the square root $\sqrt{d_i}$ of each node must be in the range $[\sqrt{d} - \gamma\sqrt{d}, \sqrt{d} + \gamma\sqrt{d}]$, so we also have the spectral bound:

$$\|D^{1/2} - \sqrt{d}I\| \leqslant \gamma\sqrt{d}. \tag{11}$$

We know that $\|D\| \leqslant d + \gamma d < 2d$ and that $\|N\| = 1$, so from (11) we get

$$\begin{aligned}
\|A - dN\| &= \|D^{1/2}ND^{1/2} - dN\| \\
&\leqslant \|D^{1/2}ND^{1/2} - \sqrt{d}ND^{1/2}\| \\
&\quad + \|\sqrt{d}ND^{1/2} - dN\| \\
&= \|(D^{1/2} - \sqrt{d}I) \cdot ND^{1/2}\| \\
&\quad + \|\sqrt{d}N \cdot (D^{1/2} - \sqrt{d}I)\| \\
&\leqslant \|D^{1/2} - \sqrt{d}I\| \cdot \|N\| \cdot \|D^{1/2}\| \\
&\quad + \sqrt{d} \cdot \|N\| \cdot \|D^{1/2} - \sqrt{d}I\| \leqslant 3\gamma d. \qquad (12)
\end{aligned}$$

By using the triangle inequality and (12) we get

$$\|N - (1/d)B\| \leqslant \|N - (1/d)A\| + (1/d) \cdot \|A - B\| \leqslant 4\gamma. \tag{13}$$

Finally, we use Theorem 19 (see Section 4.1), which is a standard fact in matrix approximation theory: if two real symmetric matrices are close in spectral norm then their eigenvalues are close. From (13) and the fact that all eigenvalues of $(1/d)B$ except for the first and second one are 0, for each $i \in \{3, \ldots, 2n\}$ we have

$$|\lambda_i| = |\lambda_i - 0| \leqslant \|N - \frac{1}{d}B\| \leqslant 4\gamma. \tag{14}$$

Similarly, from the fact that the second eigenvalue of $(1/d)B$ is $1 - 2b/d$ we get

$$|\lambda_2 - (1 - 2b/d)| \leqslant \|N - \frac{1}{d}B\| \leqslant 4\gamma,$$

that is, from hypothesis $\nu > 12\gamma$ and (14), $\lambda_2 \geqslant (1 + \delta)\lambda_3$ for some constant $\delta > 0$. This concludes the proofs of Lemma 9 and Theorem 7. ▣

By combining Lemma 7 and Theorem 3, we achieve weak reconstruction for the stochastic block model.

COROLLARY 2 (Reconstruction in Stochastic Block Models). *Let $G \sim \mathcal{G}_{2n,p,q}$. If $a - b > 25\sqrt{d \log n}$ and $b = \Omega(\log n/n^2)$ then the Averaging protocol produces an $\mathcal{O}(d \log n/(a-b)^2)$-weak reconstruction in $\mathcal{O}(\log n)$ rounds w.h.p.*

SKETCH OF PROOF. From Lemma 7 we get that w.h.p. G is $(2n, d, b, \gamma)$-clustered with

- $\gamma \leqslant 6\sqrt{\log n/d}$,
- $|\lambda_i| \leqslant 4\gamma$ for all $i = 3, \dots, 2n$ and
- $\lambda_2 \geqslant (1+\delta)\lambda_3$ for some constant $\delta > 0$.

Given the hypotheses on a and b, we also have that the graph is connected, w.h.p. Moreover, since $d\nu = (a-b) > 25\sqrt{d \log n}$, then

$$\frac{\gamma}{\nu - \lambda_3} = \frac{d\gamma}{d\nu - d\lambda_3} \leqslant \frac{6\sqrt{d\log n}}{(a-b) - 24\sqrt{d\log n}} = \mathcal{O}\left(\frac{\sqrt{d\log n}}{(a-b)}\right).$$

Theorem 3 then guarantees that the Averaging protocol finds an $\mathcal{O}(d \log n /(a-b)^2)$-weak reconstruction, w.h.p. ▣

4.6. Technical Proofs for Clustered Graphs

LEMMA 10. *Let G be a connected $(2n, d, b, \gamma)$-clustered graph (see Definition 5) with $\gamma \leqslant 1/10$. If $\lambda_3 < \nu$ then*

$$\lambda_2 \geqslant \nu - 10\gamma \qquad \text{and} \qquad \left\| D^{1/2}\boldsymbol{\chi} - \beta_2 \mathbf{w}_2 \right\| \leqslant \frac{44\,\gamma}{\nu - \lambda_3}\sqrt{2nd},$$

where $\beta_2 = \boldsymbol{\chi}^\intercal D^{1/2}\mathbf{w}_2$.

PROOF. For every node v, let us name a_v and b_v the numbers of neighbors of v in its own cluster and in the other cluster, respectively, and $d_v = a_v + b_v$ its degree. Since from the definition of $(2n, d, b, \gamma)$-clustered graph it holds that $(1-\gamma)d \leqslant d_v \leqslant (1+\gamma)d$ and $b - \gamma d \leqslant b_v \leqslant b + \gamma d$, it is easy to check that

$$|a_v - b_v - \nu d_v| \leqslant 4d\,\gamma$$

for any node v. Hence,

$$\begin{aligned}
\|A\boldsymbol{\chi} - \nu D\boldsymbol{\chi}\|^2 &= \sum_{v\in[2n]} \left(\sum_{w \in \mathrm{Neigh}(v)} \boldsymbol{\chi}(w) - \nu d_v \boldsymbol{\chi}(v) \right)^2 \\
&= \sum_{v\in[2n]} \left(a_v\boldsymbol{\chi}(v) - b_v\boldsymbol{\chi}(v) - \nu d_v \boldsymbol{\chi}(v) \right)^2 \\
&= \sum_{v\in[2n]} (a_v - b_v - \nu d_v)^2 \leqslant 32nd^2\gamma^2.
\end{aligned}$$

Thus,

$$\begin{aligned}\left\|ND^{1/2}\boldsymbol{\chi}-\nu D^{1/2}\boldsymbol{\chi}\right\| &= \left\|D^{-1/2}A\boldsymbol{\chi}-\nu D^{1/2}\boldsymbol{\chi}\right\| \\ &= \left\|D^{-1/2}\left(A\boldsymbol{\chi}-\nu D\boldsymbol{\chi}\right)\right\| \\ &\leqslant \left\|D^{-1/2}\right\|\cdot\|A\boldsymbol{\chi}-\nu D\boldsymbol{\chi}\| \\ &\leqslant \frac{2}{\sqrt{d}}\cdot\sqrt{2n}4d\,\gamma = 8\sqrt{2nd}\,\gamma.\end{aligned} \tag{15}$$

Observe that $\mathbf{w}_1$ is parallel to $D^{1/2}\mathbf{1}$ and we have that

$$|\mathbf{1}^\intercal D\boldsymbol{\chi}| = \left|\sum_{v\in[2n]}\boldsymbol{\chi}(v)d_v\right| \leqslant (1+\gamma)dn-(1-\gamma)dn = 2nd\,\gamma. \tag{16}$$

Hence, if we name $\mathbf{y}$ the component of $D^{1/2}\boldsymbol{\chi}$ orthogonal to the first eigenvector, we can write it as

$$D^{1/2}\boldsymbol{\chi} = \frac{\mathbf{1}^\intercal D\boldsymbol{\chi}}{\|D^{1/2}\mathbf{1}\|^2}D^{1/2}\mathbf{1}+\mathbf{y}. \tag{17}$$

Thus,

$$\begin{aligned}\|N\mathbf{y}-\nu\mathbf{y}\| &= \left\|N\left(D^{1/2}\boldsymbol{\chi}-\frac{\mathbf{1}^\intercal D\boldsymbol{\chi}}{\|D^{1/2}\mathbf{1}\|^2}D^{1/2}\mathbf{1}\right)\right. \\ &\qquad \left.-\nu\left(D^{1/2}\boldsymbol{\chi}-\frac{\mathbf{1}^\intercal D\boldsymbol{\chi}}{\|D^{1/2}\mathbf{1}\|^2}D^{1/2}\mathbf{1}\right)\right\| \\ &\leqslant \left\|ND^{1/2}\boldsymbol{\chi}-\nu D^{1/2}\boldsymbol{\chi}\right\| \\ &\qquad +\frac{|\mathbf{1}^\intercal D\boldsymbol{\chi}|}{\|D^{1/2}\mathbf{1}\|^2}\left\|ND^{1/2}\mathbf{1}-\nu D^{1/2}\mathbf{1}\right\| \\ &= \left\|ND^{1/2}\boldsymbol{\chi}-\nu D^{1/2}\boldsymbol{\chi}\right\|+\frac{|\mathbf{1}^\intercal D\boldsymbol{\chi}|}{\|D^{1/2}\mathbf{1}\|}\,\frac{2b}{d} \\ &\leqslant 8\sqrt{2nd}\,\gamma+4\sqrt{2nd}\,\gamma,\end{aligned} \tag{18}$$

where in the last inequality we used (15) and (16) and the facts that $b\leqslant d/2$ and $\left\|D^{1/2}\mathbf{1}\right\|\geqslant(1/2)\sqrt{2nd}$. From (17) it follows that

$$\begin{aligned}\|\mathbf{y}\| &\geqslant \left\|D^{1/2}\boldsymbol{\chi}\right\|-\frac{\mathbf{1}^\intercal D\boldsymbol{\chi}}{\left\|D^{1/2}\mathbf{1}\right\|} \\ &\geqslant (1-\gamma)\sqrt{2nd}-4\gamma\sqrt{2nd} \\ &= (1-5\gamma)\sqrt{2nd}\geqslant(1/2)\sqrt{2nd}.\end{aligned} \tag{19}$$

Now, let us we write $\mathbf{y}$ as a linear combination of the orthonormal eigenvectors of N, $\mathbf{y}=\beta_2\mathbf{w}_2+\cdots+\beta_n\mathbf{w}_n$ (recall that $\mathbf{y}^\intercal\mathbf{w}_1=0$ by definition of

$\mathbf{y}$ in (17)). From (18) and (19), it follows that

$$100\gamma^2\|\mathbf{y}\|^2 \geqslant \|N\mathbf{y}-\nu\mathbf{y}\|^2 = \left\|\sum_{i=2}^{n}(\lambda_i-\nu)\beta_i\mathbf{w}_i\right\|^2 = \sum_{i=2}^{n}(\lambda_i-\nu)^2\beta_i^2. \tag{20}$$

Moreover, from hypothesis $\lambda_3 < \nu$ we have that

$$\begin{aligned}\sum_{i=2}^{n}(\lambda_i-\nu)^2\beta_i^2 &\geqslant \sum_{i=3}^{n}(\lambda_i-\nu)^2\beta_i^2\\ &\geqslant (\lambda_3-\nu)^2\sum_{i=3}^{n}\beta_i^2\\ &= (\lambda_3-\nu)^2\|\mathbf{y}-\beta_2\mathbf{w}_2\|^2.\end{aligned} \tag{21}$$

Thus, by combining together (20) and (21) we get

$$\|\mathbf{y}-\beta_2\mathbf{w}_2\| \leqslant \frac{10\,\gamma}{\nu-\lambda_3}\|\mathbf{y}\|$$

where $\beta_2 = \mathbf{y}^\intercal\mathbf{w}_2 = \left(D^{1/2}\boldsymbol{\chi}\right)^\intercal\mathbf{w}_2$.

As for the first thesis of the lemma, observe that if $\lambda_2 \geqslant \nu$ then the first thesis is obvious. Otherwise, if $\lambda_2 < \nu$, then $(\lambda_2-\nu)^2 \leqslant (\lambda_3-\nu)^2 \leqslant \cdots \leqslant (\lambda_n-\nu)^2$. Thus, the first thesis follows from (20) and the fact that

$$\sum_{i=2}^{n}(\lambda_i-\nu)^2\beta_i^2 \geqslant (\lambda_2-\nu)^2\sum_{i=2}^{n}\beta_i^2 = (\lambda_2-\nu)^2\|\mathbf{y}\|^2.$$

As for the second thesis of the lemma, we have

$$\begin{aligned}\left\|D^{1/2}\boldsymbol{\chi}-\beta_2\mathbf{w}_2\right\| &= \left\|\frac{\mathbf{1}^\intercal D\boldsymbol{\chi}}{\|D^{1/2}\mathbf{1}\|^2}D^{1/2}\mathbf{1}+\mathbf{y}-\beta_2\mathbf{w}_2\right\|\\ &\leqslant \frac{|\mathbf{1}^\intercal D\boldsymbol{\chi}|}{\|D^{1/2}\mathbf{1}\|}+\|\mathbf{y}-\beta_2\mathbf{w}_2\|\\ &\leqslant 4\,\gamma\sqrt{2nd}+\frac{10\,\gamma}{\nu-\lambda_3}\|\mathbf{y}\|\\ &\leqslant 4\,\gamma\sqrt{2nd}+\frac{20\,\gamma}{\nu-\lambda_3}\sqrt{2nd} \leqslant \frac{44\,\gamma}{\nu-\lambda_3}\sqrt{2nd},\end{aligned}$$

where in the last inequality we used that $\mathbf{y}$ is the projection of $D^{\frac{1}{2}}\boldsymbol{\chi}$ on $D^{\frac{1}{2}}\mathbf{1}$, and thus $\|\mathbf{y}\| \leqslant \|D^{\frac{1}{2}}\boldsymbol{\chi}\| \leqslant 2\sqrt{2nd}$. ▣

4.7. Tight Analysis for the Stochastic Block Model

In Lemma 7 we have shown that, when $(a-b) > \sqrt{(a+b)\log n}$, a graph sampled according to $\mathcal{G}_{2n,p,q}$ satisfies the hypothesis of Theorem 3, w.h.p. The simple Averaging protocol thus gets weak-reconstruction in $\mathcal{O}(\log n)$ rounds. As for the parameters' range of $\mathcal{G}_{2n,p,q}$, we know that the above result is still off by a factor $\sqrt{\log n}$ from the threshold $(a-b) > 2\sqrt{(a+b)}$ [**MNS13, Mas14, MNS14**], the latter being a necessary condition for any

(centralized or not) non-trivial weak reconstruction. Essentially, the reason behind this gap is that, while Theorem 3 holds for *any* (i.e. "worst-case") $(2n, d, b, \gamma)$-clustered graph, in order to apply it to $\mathcal{G}_{2n,p,q}$ we need to choose parameters a and b in a way that γd bounds the variation of the degree of *any* node w.r.t. the regular case, w.h.p.

On the other hand, since the degrees in $\mathcal{G}_{2n,p,q}$ are distributed according to a sum of Bernoulli random variables, the rare event that some degrees are much higher than the average does not affect too much the eigenvalues and eigenvectors of the graph. Indeed, by adopting ad-hoc arguments for $\mathcal{G}_{2n,p,q}$, we prove that the Averaging protocol actually achieves an $\mathcal{O}(d/(a-b)^2)$-weak reconstruction, w.h.p., provided that

$$(a-b)^2 > c_{\text{opt}}(a+b) > 5\log n,$$

thus matching the weak-reconstruction threshold up to a constant factor for graphs of logarithmic degree. The main argument relies on the spectral properties of $\mathcal{G}_{2n,p,q}$ stated in the following lemma, whose complete proof is given in Section 4.9.

LEMMA 11. *Let $G \sim \mathcal{G}_{2n,p,q}$. If*

$$(a-b)^2 > c_{\text{opt}}(a+b) > 5\log n,$$

and[5] $a+b < n^{\frac{1}{3}-c_{\text{tight}}}$ *for some positive constants c_{opt} and c_{tight}, then w.h.p.*

(1) for some constant $c_{\text{eigerr}} > 0$, it holds

$$\lambda_2 \geqslant 1 - 2b/d - c_{\text{eigerr}}/\sqrt{d},$$

(2) $\lambda_2 \geqslant (1+\delta)\lambda$ for some constant $\delta > 0$ (where as usual $\lambda = \max\{|\lambda_3|, \ldots, |\lambda_{2n}|\}$),

(3) for each $i \in V \setminus S$, for some subset S with $|S| = \mathcal{O}(nd/(a-b)^2)$, it holds

$$|\sqrt{2nd}(D^{-1/2}\mathbf{w}_2)(i) - \boldsymbol{\chi}(i)| \leqslant \frac{1}{100}.$$

IDEA OF PROOF. The key-steps of the proof are two concentration of probability results.

In Lemma 15, we prove a tight bound on the deviation of the Laplacian $\mathcal{L}(A) = I - N$ of $\mathcal{G}_{2n,p,q}$ from the Laplacian of the expected matrix $\mathcal{L}(B) = I - \frac{1}{d}B$. As one may expect from previous results on the Erdős-Rényi model and from Le and Vershynin's concentration results for inhomogeneous Erdős-Rényi graph (see Lemma 14), we can prove that w.h.p.

$$\|\mathcal{L}(A) - \mathcal{L}(B)\| = \mathcal{O}(\sqrt{d}),$$

even when $d = \Theta(\log n)$. To derive the latter result, we leverage on the aforementioned Le and Vershynin's bound on the spectral norm of inhomogeneous Erdős-Rényi graphs; in $\mathcal{G}_{2n,p,q}$ this bound implies that if $d = \Omega(\log n)$ then w.h.p. $\|A - B\| = \mathcal{O}(\sqrt{d})$. Then, while Le and Vershynin replace the Laplacian matrix with regularized versions of it, we are able to bound

$\|\mathcal{L}(A) - \mathcal{L}(B)\|$ directly by upper bounding it with $\|A - B\|$ and an additional factor $\|B - d^{-1} D^{1/2} B D^{1/2}\|$. We then bound from above the latter additional factor thanks to our second result: In Lemma 16 (whose proof can be found at the end of the chapter), we prove that w.h.p.

$$\sum(\sqrt{d_i} - \sqrt{d})^2 \leqslant 2n \quad \text{and} \quad \sum(d_i - d)^2 \leqslant 2nd.$$

We can then prove the first two claims of Lemma 11 by bounding the distance of the eigenvalues of N from those of $d^{-1} B$ via Lemma 19.

As for the third claim of the lemma, we prove it by upper bounding the components of $D^{-1/2}\mathbf{w}$ orthogonal to $\boldsymbol{\chi}$. In particular, we can limit the projection $\mathbf{w_1}$ of $D^{-1/2}\mathbf{w}$ on $\mathbf{1}$ by using Lemma 16. Then, we can upper bound the projection $\mathbf{w}_\perp$ of $D^{-1/2}\mathbf{w}$ on the space orthogonal to both $\boldsymbol{\chi}$ and $\mathbf{1}$ with Lemma 15: We look at N as a perturbed version of B and apply the Davis-Kahan theorem. Finally, we conclude the proof observing that

$$\left\| \mathbf{w}_2 - \frac{1}{\sqrt{2n}} \right\| \leqslant 2(\|\mathbf{w_1}\| + \|\mathbf{w}_\perp\|).$$

gg

Once we have Lemma 11 we can prove the main theorem on $\mathcal{G}_{2n,p,q}$ with the same argument used for Theorem 3 (the full proof is given in Section 4.9).

THEOREM 4 (Tight Reconstruction in Stochastic Block Models). *Let $G \sim \mathcal{G}_{2n,p,q}$. If*

$$(a-b)^2 > c_{\text{opt}}(a+b) > 5 \log n,$$

and[6] *$a + b < n^{\frac{1}{3} - c_{\text{tight}}}$ for some positive constants c_{opt} and c_{tight}, then the Averaging protocol produces an $\mathcal{O}(d/(a-b)^2)$-weak reconstruction within $\mathcal{O}(\log n)$ rounds w.h.p.*

PROOF. For any vector $\mathbf{x}$, we can write

$$\mathbf{x}^{(t)} = P^t \mathbf{x} = \sum_{i=1}^{2n} a_i \lambda_i^t D^{-1/2} \mathbf{w}_i = \alpha_1 \mathbf{1} + a_2 \lambda_2^t D^{-1/2} \mathbf{w}_2 + \mathbf{e}^{(t)},$$

where $\alpha_1 = \frac{\mathbf{1}^\intercal D \mathbf{x}}{\|D^{1/2}\mathbf{1}\|}$ and $\|\mathbf{e}^{(t)}\| \leqslant 4\lambda^t \|\mathbf{x}\|$.

From Lemma 11 (Claim 3) we have that for at least $2n - \mathcal{O}(nd/(a-b)^2)$ entries i of $D^{-1/2}\mathbf{w}_2$, we get

$$|\sqrt{2nd}(D^{-1/2}\mathbf{w}_2)(i) - \boldsymbol{\chi}(i)| \leqslant \frac{1}{100},$$

[6] It should be possible to weaken the condition $d < n^{\frac{1}{3} - c_{\text{tight}}}$ via some stronger concentration argument; see the proof of Lemma 16 at the end of the chapter for details.

that is

$$(D^{-1/2}\mathbf{w}_2)(i) \geqslant \frac{99}{100\sqrt{2nd}} \quad \text{if } i \in V_1 \cap S \text{ and}$$
$$(D^{-1/2}\mathbf{w}_2)(i) \leqslant -\frac{99}{100\sqrt{2nd}} \quad \text{if } i \in V_2 \cap S.$$

Thus, we get

$$\begin{aligned}\left|\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)}\right| &= \left|a_2\lambda_2^{t-1}(\lambda_2 - 1)D^{-1/2}\mathbf{w}_2 + \mathbf{e}^{(t)} + \mathbf{e}^{(t-1)}\right| \\ &\leqslant \left|a_2\lambda_2^{t-1}(\lambda_2 - 1)D^{-1/2}\mathbf{w}_2\right| + \left|\mathbf{e}^{(t)} - \mathbf{e}^{(t-1)}\right| \qquad (22)\end{aligned}$$

and, when

$$t - 1 \geqslant \frac{\log\left(\frac{16\sqrt{2n}}{|a_2|(1-\lambda_2)}\right)}{\log\left(\frac{\lambda_2}{\lambda}\right)},$$

from (22) it follows that

$$(\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)})(i) \geqslant \frac{99}{200\sqrt{2nd}}\, a_2\lambda_2^{t-1}(\lambda_2 - 1) \quad \text{if } i \in V_j \cap S \text{ and}$$
$$(\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)})(i) \leqslant -\frac{99}{200\sqrt{2nd}}\, a_2\lambda_2^{t-1}(\lambda_2 - 1) \quad \text{if } i \in V_{3-j} \cap S.$$

either for $j = 1$ or for $j = 2$. Since

$$|S| > n - \mathcal{O}\left(\frac{nd}{(a-b)^2}\right),$$

we thus get a $\mathcal{O}(d/(a-b)^2)$-weak reconstruction. ∎

4.8. Moving Beyond Two Communities: An Outlook

The Averaging protocol can be naturally extended to address the case of more communities. One way to achieve this is by performing a suitable number of independent, parallel runs of the protocol. We next outline the analysis for a natural generalization of the regular block model. This allows us to easily present the main ideas and to provide an intuition of how and why the protocol works.

Let $G = (V, E)$ be a d-regular graph in which V is partitioned into k equally-sized communities $V_1, \ldots, V_k$, while every node in V_i has exactly a neighbors within V_i and exactly b neighbors in each V_j, for $j \neq i$. Note that

$$d = a + (k-1) \cdot b.$$

It is easy to see that the transition matrix P of the random walk on G has an eigenvalue $(a-b)/d$ with multiplicity $k-1$. The eigenspace of $(a-b)/d$ consists of all stepwise vectors that are constant within each community V_i and whose entries sum to zero. If

$$\max\{|\lambda_{2n}|, \lambda_{k+1}\} < (1-\varepsilon) \cdot \frac{a-b}{d},$$

P has eigenvalues $\lambda_1 = 1$ and

$$\lambda_2 = \cdots = \lambda_k = \frac{a-b}{d},$$

with all other eigenvalues strictly smaller by a $(1-\varepsilon)$ factor.

Let T be a large enough threshold such that, for all $t \geqslant T$, $\lambda_2^t > n^2 \lambda_{k+1}^t$ and note that T is in the order of $(1/\varepsilon)\log n$. Let $\mathbf{x} \in \mathbb{R}^V$ be a vector. We say that a node v is of *negative type* with respect to $\mathbf{x}$ if, for all $t > T$, the value $(P^t\mathbf{x})_v$ decreases with t. We say that a node v is of *positive type* with respect to $\mathbf{x}$ if, for all $t > T$, the value $(P^t\mathbf{x})_v$ increases with t. Note that a node might have neither type, because $(P^t\mathbf{x})_v$ might not be strictly monotone in t for all $t > T$.

We prove the following: If we pick ℓ random vectors $\mathbf{x}^1, \ldots, \mathbf{x}^\ell$, each in $\{-1,1\}^V$, then w.h.p.

i) every node is either of positive or negative type for each $\mathbf{x}^i$[7];
ii) furthermore, if we associate a "signature" to each node, namely, the sequence of ℓ types, then nodes within the same V_i exhibit the same signature, while nodes in different V_i, V_j have different signatures.

These are the basic intuitions that allow us to prove the following theorem.

THEOREM 2 (More Communities). *Let $G = (V, E)$ be a k-clustered d-regular graph defined as above and assume that*

$$\lambda = \max\{|\lambda_{2n}|, \lambda_{k+1}\} < (1-\varepsilon) \cdot \frac{a-b}{d},$$

for a suitable constant $\varepsilon > 0$. Then, for $\ell = \Theta(\log n)$, the Averaging protocol with ℓ parallel runs produces a strong reconstruction within $\mathcal{O}(\log n)$ rounds, w.h.p.

The proof is divided in the following two lemmas.

LEMMA 12. *Pick $\mathbf{x} \sim \{-1,1\}^{kn}$ u.a.r. Then the nodes of V_1 are either all of positive type or all of negative type, w.h.p. Furthermore, the two events have equal probability.*

PROOF. We write

$$\mathbf{x} = \mathbf{x_1} + \mathbf{x}_{V_1} + \mathbf{x}_{\perp_1} + \mathbf{x}_\perp,$$

where $\mathbf{x_1}$ is the component of $\mathbf{x}$ parallel to $\mathbf{1}$, $\mathbf{x}_{V_1}$ is the component parallel to the vector $\mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V$, $\mathbf{x}_{\perp_1}$ is the component in the eigenspace of λ_2 and orthogonal to $\mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V$, and $\mathbf{x}_\perp$ is the component orthogonal to $\mathbf{1}$ and to the eigenspace of λ_2.

For the above the make sense, $\mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V$ must be an eigenvector of λ_2, which is easily verified because its entries sum to zero and they are constant within components.

[7] I.e., for every $t > T$, $(P^t\mathbf{x})_v$ monotonically increases (or decreases) w.r.t. t.

An important observation, and the reason for picking the above decomposition, is that $\mathbf{x}_{\perp_1}$ is zero in V_1. The reason is that $\mathbf{x}_{\perp_1}$ has to be orthogonal to $\mathbf{1}_V$ and to $\mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V$ so from

$$\langle \mathbf{x}_{\perp_1}, \mathbf{1}_V \rangle = \langle \mathbf{x}_{\perp_1}, \mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V \rangle = 0,$$

we deduce

$$\langle \mathbf{x}_{\perp_1}, \mathbf{1}_{V_1} \rangle = 0.$$

Thus, the entries of $\mathbf{x}_{\perp_1}$ sum to zero within V_1, but, being in the eigenspace of λ_2, the entries of $\mathbf{x}_{\perp_1}$ are constant within components, and so they must be all zero within V_1.

Now we have

$$P^t\mathbf{x} = \mathbf{x_1} + \lambda_2^t \mathbf{x}_{V_1} + \lambda_2^t \mathbf{x}_{\perp_1} + P^t\mathbf{x}_\perp,$$

and so, for each $v \in V_1$ it holds

$$(P^{t+1}\mathbf{x})_v - (P^t\mathbf{x})_v = \lambda_2^t \cdot (1-\lambda_2)(\mathbf{x}_{V_1})_v + ((P^{t+1} - P^t)\mathbf{x}_\perp)_v. \tag{23}$$

For $t > T$, the hypothesis $\lambda < (1-\varepsilon)\lambda_2$ implies that

$$\begin{aligned} |(P^t\mathbf{x}_\perp)_v| &\leqslant ||P^t\mathbf{x}_\perp||_\infty \leqslant ||P^t\mathbf{x}_\perp|| \\ &\leqslant \lambda^t ||\mathbf{x}_\perp|| \leqslant \sqrt{n}\cdot\lambda^t \leqslant \frac{1}{n^{1.5}}\lambda_2^t. \end{aligned} \tag{24}$$

Moreover, for each $v \in V_1$ we have

$$\begin{aligned} |(\mathbf{x}_{V_1})_v| &= \|\mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V\|^{-2}\langle \mathbf{x}, \mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V\rangle \left(1 - k^{-1}\right) \\ &= \frac{k}{(k-1)n}\left(\sum_{i\in V_1} x_i - \sum_{i\in V}\frac{x_i}{k}\right)\left(\frac{k-1}{k}\right) \\ &= \frac{1}{n}\left(\sum_{i\in V_1} x_i - \sum_{i\in V}\frac{x_i}{k}\right), \end{aligned}$$

and

$$||\mathbf{x}_{V_1}|| = \frac{\langle \mathbf{x}, \mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V\rangle}{\|\mathbf{1}_{V_1} - k^{-1}\mathbf{1}_V\|} = \sqrt{\frac{k}{(k-1)n}}\cdot\left(\sum_{i\in V_1} x_i - \sum_{i\in V}\frac{x_i}{k}\right),$$

which imply that

$$|(\mathbf{x}_{V_1})_v| = \sqrt{\frac{1-1/k}{n}}\cdot\|\mathbf{x}_{V_1}\|. \tag{25}$$

Finally, note that by Lemma 3 it holds w.h.p. $||\mathbf{x}_{V_1}|| \geqslant \frac{1}{n}||\mathbf{x}|| \geqslant \sqrt{k/n}$.

The latter fact together with (24) and (25) imply that w.h.p. the sign of (23) is the same as the sign of $(\mathbf{x}_{V_1})_v$, which is the same for all elements of V_1 and is equally likely to be positive or negative. ∎

Of course the same statement is true if we replace V_1 by V_i for any $i = 1, \ldots, k$; by a union bound, it is also true for all i simultaneously, w.h.p.

LEMMA 13. *Pick $\mathbf{x} \sim \{-1,1\}^{kn}$ u.a.r. There is an absolute constant p (e.g., $p = \frac{1}{100}$) such that, with probability at least p, all nodes of V_1 have the same type, all nodes of V_2 have the same type, and the types are different.*

PROOF. This time we write

$$\mathbf{x} = \mathbf{x_1} + \mathbf{x}_{V_{1plus2}} + \mathbf{x}_{V_{1minus2}} + \mathbf{x}_{\perp_{1,2}} + \mathbf{x}_{\perp}$$

where

- $\mathbf{x_1}$ is the component parallel to $\mathbf{1}_V$,
- $\mathbf{x}_{V_{1plus2}}$ is the component parallel to $\mathbf{1}_{V_1} + \mathbf{1}_{V_2} - \frac{2}{k}\mathbf{1}_V$,
- $\mathbf{x}_{V_{1minus2}}$ is the component parallel to $\mathbf{1}_{V_1} - \mathbf{1}_{V_2}$,
- $\mathbf{x}_{\perp_{1,2}}$ is the component in the eigenspace of λ_2 and orthogonal to $\mathbf{x}_{V_{1plus2}}$ and $\mathbf{x}_{V_{1minus2}}$,
- $\mathbf{x}_{\perp}$ is the rest.

Similarly to the proof of Lemma 12, the important observations are that $\mathbf{x}_{V_{1plus2}}$ and $\mathbf{x}_{V_{1minus2}}$ are in the eigenspace of λ_2, and that $\mathbf{x}_{\perp_{1,2}}$ is zero in all the coordinates of V_1 and of V_2.

Thus, for each $v \in V_1 \cup V_2$ we have

(26)
$$(P^{t+1}\mathbf{x})_v - (P^t\mathbf{x})_v = \lambda_2^t(1-\lambda_2)(\mathbf{x}_{V_{1plus2}} + \mathbf{x}_{V_{1minus2}})_v + ((P^{t+1} - P^t)\mathbf{x}_{\perp})_v.$$

From (26) it is easy to see that if $\mathbf{x}$ is such that, for every $v \in V_1 \cup V_2$, we have the two conditions

$$|(\mathbf{x}_{V_{1plus2}})_v| \leqslant \frac{3}{4}|(\mathbf{x}_{V_{1minus2}})_v| \quad \text{and} \tag{27}$$

$$|((P^{t+1} - P^t)\mathbf{x}_{\perp})_v| \leqslant \frac{1}{8}\lambda_2^t \cdot (1-\lambda_2) \cdot |(\mathbf{x}_{V_{1minus2}})_v|, \tag{28}$$

then such an $\mathbf{x}$ satisfies the conditions of the Lemma, that is all the elements in V_1 have the same type, all the elements of V_2 have the same type, and the types are different. Now note that, since

$$|(\mathbf{x}_{V_{1plus2}})_v| = \frac{1}{2n}\left(\sum_{i\in V_1} x_i + \sum_{i\in V_1} x_i - \frac{2}{k}\sum_{i\in V} x_i\right) \quad \text{and}$$

$$|(\mathbf{x}_{V_{1minus2}})_v| = \frac{1}{2n}\left(\sum_{i\in V_1} x_i - \sum_{i\in V_2} x_i\right),$$

if $\mathbf{x}$ satisfies

$$2\sqrt{n} \leqslant \sum_{v\in V_1} x_v \leqslant 3\sqrt{n}, \tag{29}$$

$$-2\sqrt{n} \leqslant \sum_{v\in V_2} x_v \leqslant -\sqrt{n} \quad \text{and} \tag{30}$$

$$0 \leqslant \sum_{v\in V/(V_1\cup V_2)} x_v \leqslant \frac{1}{10}\sqrt{kn}, \tag{31}$$

then (27) is satisfied, and note that (29), (30) and (31) are independent and each happens with constant probability.

Finally, observe that if (27) holds then (28) is satisfied with high probability when $t > T$. ◻

It is enough to pick $\ell = \log(3n)$ to have that the signatures are well defined and they are the same within each community and different between communities, w.h.p. The first lemma guarantees that, for all ℓ vectors, all nodes within each community have the same type, w.h.p. The second lemma guarantees that the signatures are different between communities, w.h.p.

4.9. Technical Proofs for Stochastic Block Models

LEMMA 11. *Let $G \sim \mathcal{G}_{2n,p,q}$. If*

$$(a-b)^2 > c_{\text{opt}}(a+b) > 5\log n,$$

and[7] *$a+b < n^{\frac{1}{3}-c_{\text{tight}}}$ for some positive constants c_{opt} and c_{tight}, then w.h.p.*

1. *for some constant $c_{\text{eigerr}} > 0$, it holds*
$$\lambda_2 \geqslant 1 - 2b/d - c_{\text{eigerr}}/\sqrt{d},$$
2. *$\lambda_2 \geqslant (1+\delta)\lambda$ for some constant $\delta > 0$ (where as usual $\lambda = \max\{|\lambda_3|, \dots, |\lambda_{2n}|\}$),*
3. *for each $i \in V \setminus S$, for some subset S with $|S| = \mathcal{O}(nd/(a-b)^2)$, it holds*
$$|\sqrt{2nd}(D^{-1/2}\mathbf{w}_2)(i) - \boldsymbol{\chi}(i)| \leqslant \frac{1}{100}.$$

PROOF. Let G be a randomly-generated graph according to $\mathcal{G}_{2n,p,q}$ with $a = pn$, $b = qn$ and $d = a + b$. Recall the definitions of A, D, N, P, λ_i and $\mathbf{w}_i$ ($i \in \{1, \dots, 2n\}$) in Section 4.2, and let B be defined as in Section 4.5.1. Let us denote with A_i ($i \in \{1, 2\}$) the adjacency matrix of the subgraph of G induced by community V_i, with $A_B = \{A_{u,v-n}\}_{u \in V_1, v \in V_2}$ the matrix whose entry (i, j) is 1 iff there is an edge between the i-th node of V_1 and the j-th node of V_2, then

$$A = \begin{pmatrix} A_1 & A_B \\ A_B^\intercal & A_2 \end{pmatrix}.$$

We need the following technical lemmas.

LEMMA 14. *If $d > 5\log n$ then for some positive constant c_{spect} it holds w.h.p.*

$$\|A - B\| \leqslant c_{\text{spect}}\sqrt{d}.$$

PROOF OF LEMMA 14. The lemma directly follows from Theorem 2.1 in [**LV15**] with $d' = 2d$ and the observation that, from the Chernoff bounds (Lemma 76), all degrees are smaller than $2d$, w.h.p. ◻ (of Lemma 14)

LEMMA 15. *If $d > 5\log n$ then for some constant $c_{\text{NvsB}} > 0$ it holds w.h.p.*

$$\|dN - B\| \leqslant c_{\text{NvsB}}\sqrt{d}.$$

The idea for proving Lemma 15 is to use the triangle inequality to upper bound $\|dN - B\|$ in terms of $\|A - B\|$, which we can bound with Lemma 14, and $\|B - 1/dD^{1/2}BD^{1/2}\|$, which we can upper bound by bounding $\|\sqrt{d}\mathbf{1} - D^{1/2}\mathbf{1}\|$ and $\|\sqrt{d}\boldsymbol{\chi} - D^{1/2}\boldsymbol{\chi}\|$ where $\mathbf{1}$ and $\boldsymbol{\chi}$ are the eigenvector corresponding to the only two non-zero eigenvalues of B. The complete proof of Lemma 15 is deferred to Section 4.9.1. As for the required bound on

$$\|\sqrt{d}\mathbf{1} - D^{1/2}\mathbf{1}\| = \|\sqrt{d}\boldsymbol{\chi} - D^{1/2}\boldsymbol{\chi}\| = \sum_{j\in V} |\sqrt{d} - \sqrt{d_j}|^2,$$

we provide it in the following lemma, whose proof is also deferred to Section 4.9.1.

LEMMA 16. *If $5\log n < d < n^{\frac{1}{3}-c_{\text{tight}}}$ for any constant $c_{\text{tight}} > 0$, it holds w.h.p.*

$$\sum_{j\in V} |\sqrt{d} - \sqrt{d_j}|^2 \leqslant 2n \text{ and}$$
$$\sum_{j\in V} |d - d_j|^2 \leqslant 2dn.$$

By combining Lemma 15 and Theorem 19 we have $|\lambda_i - \lambda_i'| \leqslant \|N - d^{-1}B\| = \mathcal{O}(1/\sqrt{d})$, where $\lambda_1' = 1$, $\lambda_2' = 1 - 2b/d$ and $\lambda_i' = 0$ for $i \in \{3, \dots, 2n\}$ are the eigenvalues of $d^{-1}B$. This proves the first two part of Lemma 11.

As for the third part, let us write $\mathbf{w}_2 = \mathbf{w}_{\mathbf{1}} + \mathbf{w}_{\boldsymbol{\chi}} + \mathbf{w}_\perp$ where $\mathbf{w}_{\mathbf{1}}$ and $\mathbf{w}_{\boldsymbol{\chi}}$ are the projection of $\mathbf{w}_2$ on $\mathbf{1}$ and $\boldsymbol{\chi}$ respectively, and $\mathbf{w}_\perp$ is the projection of $\mathbf{w}_2$ on the space orthogonal to $\mathbf{1}$ and $\boldsymbol{\chi}$.

Observe that the only non-zero eigenvalues of $(1/d)B$ are 1 and $(a-b)/d$. Thus, from Lemma 15 and the Davis-Kahan theorem (Theorem 20) with $M_1 = N$, $M_2 = \frac{1}{d}B$, $t = \lambda_2$, $\mathbf{x} = \mathbf{w}_2$ and $\delta = \lambda_2/2$, we get

$$\|\mathbf{w}_\perp\| \leqslant \frac{4}{\lambda_2\pi}\left\|N - \frac{1}{d}B\right\| \leqslant \mathcal{O}\left(\frac{1}{\sqrt{d}\lambda_2}\right) = \mathcal{O}\left(\frac{\sqrt{d}}{a-b}\right). \tag{32}$$

As for $\mathbf{w}_{\mathbf{1}}$, we know that $\langle \mathbf{w}_2, D^{-1/2}\mathbf{1}\rangle = 0$, thus

$$\begin{aligned}\|\mathbf{w}_{\mathbf{1}}\| &= \frac{1}{\sqrt{2n}}\langle \mathbf{w}_2, \mathbf{1} - d^{-\frac{1}{2}}D^{\frac{1}{2}}\mathbf{1}\rangle \\ &\leqslant \frac{1}{\sqrt{2n}}\|\mathbf{w}_2\|\|\mathbf{1} - d^{-\frac{1}{2}}D^{\frac{1}{2}}\mathbf{1}\| \leqslant \frac{1}{\sqrt{d}},\end{aligned} \tag{33}$$

where in the last inequality we used Lemma 16.

By the law of cosines and the fact that $\sqrt{1-x} \geqslant 1-x$ for $x \in [0,1]$ we have that

$$\begin{aligned} \left\| \mathbf{w}_2 - \frac{1}{\sqrt{2n}}\boldsymbol{\chi} \right\|^2 &= \|\mathbf{w}_2\|^2 + \left\| \frac{1}{\sqrt{2n}}\boldsymbol{\chi} \right\|^2 - 2\langle \mathbf{w}_2, \frac{1}{\sqrt{2n}}\boldsymbol{\chi} \rangle \\ &= 2 - 2\|\mathbf{w}_{\boldsymbol{\chi}}\| \\ &= 2 - 2\sqrt{1 - \|\mathbf{w}_{\mathbf{1}}\|^2 + \|\mathbf{w}_{\perp}\|^2} \\ &\leqslant 2\left(\|\mathbf{w}_{\mathbf{1}}\|^2 + \|\mathbf{w}_{\perp}\|^2\right) = \mathcal{O}\left(\frac{d}{(a-b)^2}\right), \end{aligned} \tag{34}$$

where in the last inequality we used (32) and (33). (34) implies that, with the exception of a set S of at most $\mathcal{O}(nd/(a-b)^2)$ nodes, we have

$$\left| \sqrt{2n}\mathbf{w}_2(i) - \boldsymbol{\chi}(i) \right| \leqslant \frac{1}{201}, \tag{35}$$

for each $i \in V/S$. From the Chernoff bound (Lemma 76), we also have that $\sqrt{d/d_i} = 1 \pm 1/201$ w.h.p. Thus, (35) and the last fact imply that for each $i \in V/S$ it holds w.h.p.

$$\left| \sqrt{2nd} D^{-\frac{1}{2}} \mathbf{w}_2(i) - \boldsymbol{\chi}(i) \right| \leqslant \frac{1}{100},$$

concluding the proof. gg

Remark 4. After looking at Lemma 11, one may wonder whether it could be enough to generalize Definition 5 to include "quasi-$(2n, d, b, \gamma)$-clustered graph", i.e. graphs that are $(2n, d, b, \gamma)$-clustered except for a small number of nodes which may have a much higher degree. In fact, this would be rather surprising: This higher-degree nodes may connect to the other nodes in such a way that would greatly perturb the eigenvalues and eigenvectors of the graph. In $\mathcal{G}_{2n,p,q}$, besides the fact that the nodes with degree much larger than d are few, it is also crucial that they are connected in a *non-adversarial* way, i.e. randomly.

4.9.1. Technical lemmas in the proof of Lemma 11

Lemma 15. *If $d > 5\log n$ then for some constant $c_{\text{NvsB}} > 0$ it holds w.h.p.*

$$\|dN - B\| \leqslant c_{\text{NvsB}}\sqrt{d}.$$

Proof. A simple application of the Chernoff bound (Lemma 76) and the union bound shows that w.h.p.

$$\sqrt{d}\|D^{-1/2}\| \leqslant 1 + \mathcal{O}\left(\sqrt{\frac{\log n}{d}}\right), \tag{36}$$

hence

$$\begin{aligned}
\|dN - B\| &= \|(\sqrt{d}D^{-1/2})A(\sqrt{d}D^{-1/2}) - B\| \\
&\leqslant \|\sqrt{d}D^{-1/2}\| \left\|A - \frac{1}{\sqrt{d}}D^{1/2}B\frac{1}{\sqrt{d}}D^{1/2}\right\| \|\sqrt{d}D^{-1/2}\| \\
&\leqslant \left\|A - \frac{1}{d}D^{1/2}BD^{1/2}\right\| \|\sqrt{d}D^{-1/2}\|^2 \\
&\leqslant \left(\|A - B\| + \left\|B - \frac{1}{d}D^{1/2}BD^{1/2}\right\|\right)\left(1 + \mathcal{O}\left(\sqrt{\frac{\log n}{d}}\right)\right). \qquad (37)
\end{aligned}$$

Thanks to Lemma 14, it holds $\|A-B\| = \mathcal{O}(\sqrt{d})$. Hence, in order to conclude the proof, it remains to show that $\|B - d^{-1}D^{1/2}BD^{1/2}\| = \mathcal{O}(\sqrt{d})$. We do that by observing that

$$\begin{aligned}
&\left\|B - \frac{1}{d}D^{1/2}BD^{1/2}\right\| \qquad (38) \\
&\leqslant \left\|B - \frac{1}{\sqrt{d}}BD^{1/2}\right\| + \left\|\frac{1}{\sqrt{d}}BD^{1/2} - \frac{1}{d}D^{1/2}BD^{1/2}\right\|,
\end{aligned}$$

and by upper-bounding the two terms on the right hand side. The two only non-zero eigenvalues of B are $a+b$ and $a-b$, with corresponding eigenvectors $(2n)^{-1/2}\,\mathbf{1}$ and $(2n)^{-1/2}\,\boldsymbol{\chi}$, therefore we can write $B = d/(2n)\,\mathbf{1}\mathbf{1}^\intercal + (a - b)/(2n)\,\boldsymbol{\chi}\boldsymbol{\chi}^\intercal$, which implies that

$$B - \frac{1}{\sqrt{d}}BD^{1/2} = \frac{\sqrt{d}}{2n}\,\mathbf{1}\,(\sqrt{d}\,\mathbf{1} - D^{1/2}\,\mathbf{1})^\intercal + \frac{a-b}{\sqrt{d}\,2n}\,\boldsymbol{\chi}\,(\sqrt{d}\,\boldsymbol{\chi} - D^{1/2}\,\boldsymbol{\chi})^\intercal.$$

It follows that, for an arbitrary unitary vector $\mathbf{x}$ it holds

$$\begin{aligned}
&\left\|\left(B - \frac{1}{\sqrt{d}}BD^{1/2}\right)\mathbf{x}\right\| \qquad (39) \\
&\leqslant \left\|\frac{\sqrt{d}}{2n}\,\mathbf{1}\,(\sqrt{d}\,\mathbf{1} - D^{1/2}\,\mathbf{1})^\intercal\mathbf{x}\right\| \\
&\quad + \left\|\frac{a-b}{\sqrt{d}\,2n}\,\boldsymbol{\chi}\,(\sqrt{d}\,\boldsymbol{\chi} - D^{1/2}\,\boldsymbol{\chi})^\intercal\mathbf{x}\right\| \\
&= \frac{\sqrt{d}}{2n}\,\|\mathbf{1}\|\,|(\sqrt{d}\,\mathbf{1} - D^{1/2}\,\mathbf{1})^\intercal\mathbf{x}| \\
&\quad + \frac{a-b}{\sqrt{d}\,2n}\,\|\boldsymbol{\chi}\|\,|(\sqrt{d}\,\boldsymbol{\chi} - D^{1/2}\,\boldsymbol{\chi})^\intercal\mathbf{x}| \\
&\leqslant \frac{\sqrt{d}}{\sqrt{2n}}\,\left\|\sqrt{d}\,\mathbf{1} - D^{1/2}\,\mathbf{1}\right\| \cdot \|\mathbf{x}\| \\
&\quad + \frac{a-b}{\sqrt{2dn}}\,\left\|\sqrt{d}\,\boldsymbol{\chi} - D^{1/2}\,\boldsymbol{\chi}\right\| \cdot \|\mathbf{x}\| \leqslant 2\sqrt{d},
\end{aligned}$$

where we used the triangle inequality, the fact that $\|\mathbf{1}\| = \|\boldsymbol{\chi}\| = \sqrt{2n}$, the Cauchy-Schwartz inequality, Lemma 16 and $a - b < d$. As for the other term on the r.h.s. of (38), we have that w.h.p.

$$\begin{aligned}(40)\qquad &\left\| \frac{1}{\sqrt{d}} BD^{1/2} - \frac{1}{d} D^{1/2} B D^{1/2} \right\| \\ &\leqslant \left\| B - \frac{1}{\sqrt{d}} D^{1/2} B \right\| \frac{1}{\sqrt{d}} \|D^{1/2}\| \leqslant 2\sqrt{d} \left(1 + \mathcal{O}\left(\sqrt{\frac{\log n}{d}} \right) \right),\end{aligned}$$

where in the last inequality we used (36) and that for any matrix M it holds $\|M\| = \|M^\intercal\|$. Finally, (39) and (40) togeter implies the desired upper bound on (38) and thus (37), concluding the proof. ⊠

LEMMA 16. *If $5 \log n < d < n^{\frac{1}{3} - c_{\text{tight}}}$ for any constant $c_{\text{tight}} > 0$, it holds w.h.p.*

$$\begin{aligned}&\sum_{j \in V} |\sqrt{d} - \sqrt{d_j}|^2 \leqslant 2n \text{ and} \\ &\sum_{j \in V} |d - d_j|^2 \leqslant 2dn.\end{aligned}$$

PROOF. Each degree d_i has the distribution of a sum of n Bernoulli random variables of expectation p plus a sum of n Bernoulli random variables of expectation q. Thus, each d_i satisfies $\mathbb{E} d_i = d$ and $\text{Var}(d_i) \leqslant d$.

First, we consider the random variables $|d - d_j|^2$. Their expectation is $\mathbb{E}|d - d_j|^2 \leqslant d$ (the variance of the random variable d_j). Let $e_{u,v}$ is the variable that is 1 iff the edge (u, v) is included in the graph. Observe that

$$\begin{aligned}|d - d_j|^4 &= |d - \sum_{v \in V} e_{j,v}|^4 \\ &= |a - \sum_{v \in V_i} e_{j,v} + b - \sum_{v \in V_{3-i}} e_{j,v}|^4 \\ &= |a - \sum_{v \in V_i} e_{j,v}|^4 + |b - \sum_{v \in V_{3-i}} e_{j,v}|^4 \\ &\quad + 6|a - \sum_{v \in V_i} e_{j,v}|^2 \, |b - \sum_{v \in V_{3-i}} e_{j,v}|^2 \\ &\quad + 4(a - \sum_{v \in V_i} e_{j,v}) \, (b - \sum_{v \in V_{3-i}} e_{j,v})^3 \\ &\quad + 4(a - \sum_{v \in V_i} e_{j,v})^3 \, (b - \sum_{v \in V_{3-i}} e_{j,v}),\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}(a - \sum_{v \in V_i} e_{j,v})^3 (b - \sum_{v \in V_{3-i}} e_{j,v}) \\
&= \mathbb{E}(a - \sum_{v \in V_i} e_{j,v})^3 \mathbb{E}(b - \sum_{v \in V_{3-i}} e_{j,v}) = 0, \\
&\mathbb{E}(a - \sum_{v \in V_i} e_{j,v})(b - \sum_{v \in V_{3-i}} e_{j,v})^3 \\
&= \mathbb{E}(a - \sum_{v \in V_i} e_{j,v}) \mathbb{E}(b - \sum_{v \in V_{3-i}} e_{j,v})^3.
\end{aligned}$$

Hence, since the fourth central moment of a binomial with parameters n and p is $np(1-p)^4 + np^4(1-p) + 3n(n-1)p^2(1-p)^2 \leqslant 4(np)^2$, if we let $i \in \{1, 2\}$ be the index of the community of j we have that the expectation of the square of $|d - d_j|^2$ (which is the fourth central moment of d_j) is

$$\begin{aligned}
\mathbb{E}|d - d_j|^4 &= \mathbb{E}|a - \sum_{v \in V_i} e_{j,v}|^4 + \mathbb{E}|b - \sum_{v \in V_{3-i}} e_{j,v}|^4 \\
&\quad + 6\mathbb{E}|a - \sum_{v \in V_i} e_{j,v}|^2 \mathbb{E}|b - \sum_{v \in V_{3-i}} e_{j,v}|^2 \\
&\leqslant 4a^2 + 4b^2 + 6ab \leqslant 4d^2.
\end{aligned}$$

In order to apply Chebyshev's inequality, we need to bound the variance of $\sum_j |d - d_j|^2$. As for the second moment of their sum, we have

$$\begin{aligned}
\mathbb{E}[(\sum_i |d - d_j|^2)^2] &= \sum_i \mathbb{E}[|d - d_j|^4] \\
&\quad + 2 \sum_{1 \leqslant i < j \leqslant 2n} \mathbb{E}[|d - d_i|^2 \cdot |d - d_j|^2] \\
&\leqslant 8d^2 n + 2 \sum_{1 \leqslant i < j \leqslant 2n} \mathbb{E}[|d - d_i|^2 \cdot |d - d_j|^2]. \qquad (41)
\end{aligned}$$

To upper bound the terms $\mathbb{E}[|d - d_i|^2 \cdot |d - d_j|^2]$, since the stochastic dependency between d_i and d_j is due only to the edge (i, j), let us write

$$d_i = \sum_{u \in N(i)} e_{i,u} = e_{i,j} + \sum_{u \in N(i)/\{j\}} e_{i,u} = e_{i,j} + d_i^{(j)},$$

where $d_i^{(j)}$ is the sum of all the edges incident to i except for (i,j). We have

$$
\begin{aligned}
(42)\quad & |d-d_i|^2\cdot|d-d_j|^2\\
&=|d-d_i^{(j)}+e_{i,j}|^2\cdot|d-d_j^{(i)}+e_{i,j}|^2\\
&=(|d-d_i^{(j)}|^2+e_{i,j}\\
&\qquad+2e_{i,j}(d-d_i^{(j)}))(|d-d_j^{(i)}|^2+e_{i,j}+2e_{i,j}(d-d_j^{(i)}))\\
&=|d-d_i^{(j)}|^2|d-d_j^{(i)}|^2+e_{i,j}|d-d_j^{(i)}|^2+2e_{i,j}(d-d_i^{(j)})|d-d_j^{(i)}|^2\\
&+|d-d_i^{(j)}|^2e_{i,j}+e_{i,j}+2e_{i,j}(d-d_i^{(j)})\\
&+2e_{i,j}(d-d_j^{(i)})|d-d_i^{(j)}|^2+2e_{i,j}(d-d_j^{(i)})\\
&\qquad+4e_{i,j}(d-d_i^{(j)})(d-d_j^{(i)}),
\end{aligned}
$$

where we used that, since $e_{i,j}$ is an indicator variable, it holds $e_{i,j}^2=e_{i,j}$. Taking the expectation of (42) we thus get

$$
\begin{aligned}
&\mathbb{E}[|d-d_i|^2\cdot|d-d_j|^2]\\
&=\mathbb{E}[|d-d_i^{(j)}|^2|d-d_j^{(i)}|^2+e_{i,j}|d-d_j^{(i)}|^2\\
&\qquad+2e_{i,j}(d-d_i^{(j)})|d-d_j^{(i)}|^2\\
&\qquad+|d-d_i^{(j)}|^2e_{i,j}+e_{i,j}+2e_{i,j}(d-d_i^{(j)})\\
&\qquad+2e_{i,j}(d-d_j^{(i)})|d-d_i^{(j)}|^2+2e_{i,j}(d-d_j^{(i)})\\
&\qquad+4e_{i,j}(d-d_i^{(j)})(d-d_j^{(i)})]\\
&=\mathbb{E}[|d-d_i^{(j)}|^2]\mathbb{E}[|d-d_j^{(i)}|^2]+\mathbb{E}[e_{i,j}]\mathbb{E}[|d-d_j^{(i)}|^2]\\
&\qquad+2\mathbb{E}[e_{i,j}]\mathbb{E}[(d-d_i^{(j)})]\mathbb{E}[|d-d_j^{(i)}|^2]\\
&\qquad+\mathbb{E}[e_{i,j}]\mathbb{E}[|d-d_i^{(j)}|^2]+\mathbb{E}[e_{i,j}]+2\mathbb{E}[e_{i,j}]\mathbb{E}[(d-d_i^{(j)})]\\
&\qquad+2\mathbb{E}[e_{i,j}]\mathbb{E}[(d-d_j^{(i)})]\mathbb{E}[|d-d_i^{(j)}|^2]+2\mathbb{E}[e_{i,j}]\mathbb{E}[(d-d_j^{(i)})]\\
&\qquad+4\mathbb{E}[e_{i,j}]\mathbb{E}[(d-d_i^{(j)})]\mathbb{E}[(d-d_j^{(i)})]\\
&\leqslant\mathbb{E}[|d-d_i^{(j)}|^2]\mathbb{E}[|d-d_j^{(i)}|^2]+\frac{d^2}{n}+2\frac{d^3}{n^2}+\frac{d^2}{n}+\frac{d}{n}\\
&\qquad+2\frac{d^2}{n^2}+2\frac{d^3}{n^2}+2\frac{d^2}{n^2}+4\frac{d^3}{n^3}\\
(43)\quad&\leqslant\mathbb{E}[|d-d_i^{(j)}|^2]\mathbb{E}[|d-d_j^{(i)}|^2]+15\frac{d^2}{n},
\end{aligned}
$$

where in the inequalities we used that $\mathbb{E}[e_{i,j}]\leqslant d/n$, that

$$
\mathbb{E}[d-d_i^{(j)}]\leqslant\mathbb{E}[e_{i,j}]+\mathbb{E}[\sum_{u\in N(i)/\{j\}}\mathbb{E}[e_{i,u}]-d_i^{(j)}]\leqslant\frac{d}{n},
$$

and that

$$\mathbb{E}[|d - d_i^{(j)}|^2] \leqslant \mathbb{E}[e_{i,j}] + \mathbb{E}[|d - \mathbb{E}[e_{i,j}] - d_i^{(j)}|^2] \leqslant \frac{d}{n} + d - 1 \leqslant d. \tag{44}$$

By combining (41) and (43) we get

$$\begin{aligned}\mathbb{E}[(\sum_i |d - d_j|^2)^2] \\ \leqslant 8d^2n + 2\sum_{1\leqslant i<j\leqslant 2n} \mathbb{E}[|d - d_i^{(j)}|^2]\mathbb{E}[|d - d_j^{(i)}|^2] + 60d^2n,\end{aligned} \tag{45}$$

As for the square of the average, we have

$$\begin{aligned}&(\mathbb{E}[\sum_i |d - d_i|^2])^2 \\ &= \sum_i \mathbb{E}[|d - d_i|^2]^2 + 2\sum_{i\neq j} \mathbb{E}[|d - d_i|^2]\mathbb{E}[|d - d_j|^2] \\ &\geqslant 2\sum_{1\leqslant i<j\leqslant 2n} \mathbb{E}[|d - d_i|^2]\mathbb{E}[|d - d_j|^2],\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}[|d - d_i|^2]\mathbb{E}[|d - d_j|^2] \\ &= \mathbb{E}[|d - d_i^{(j)} - e_{i,j}|^2]\mathbb{E}[|d - d_j^{(i)} - e_{i,j}|^2] \\ &= (\mathbb{E}[|d - d_i^{(j)}|^2] + \mathbb{E}[e_{i,j}] \\ &\qquad - 2\mathbb{E}[e_{i,j}]\mathbb{E}[(d - d_i^{(j)})]) \cdot (\mathbb{E}[|d - d_j^{(i)}|^2] \\ &\qquad + \mathbb{E}[e_{i,j}] - 2\mathbb{E}[e_{i,j}]\mathbb{E}[(d - d_j^{(i)})]) \\ &\geqslant (\mathbb{E}[|d - d_i^{(j)}|^2] - 2\mathbb{E}[e_{i,j}]\mathbb{E}[(d - d_i^{(j)})]) \cdot (\mathbb{E}[|d - d_j^{(i)}|^2] \\ &\qquad - 2\mathbb{E}[e_{i,j}]\mathbb{E}[(d - d_j^{(i)})]) \\ &\geqslant \mathbb{E}[|d - d_i^{(j)}|^2]\mathbb{E}[|d - d_j^{(i)}|^2] - 4\frac{d^3}{n^2},\end{aligned} \tag{46}$$

where we used, again, that $\mathbb{E}[e_{i,j}] \leqslant d/n$ and that $\mathbb{E}[|d - d_i^{(j)}|^2] \leqslant d$ (see (44)).

Combining (45) and (46) together we get

$$\begin{aligned}\mathrm{Var}[\sum_i |d - d_i|^2] &= \mathbb{E}[(\sum_i |d - d_i|^2)^2] - \mathbb{E}[\sum_i |d - d_i|^2]^2 \\ &\leqslant 8d^2n + 60d^2n + 16d^3 = 84d^2n\end{aligned}$$

Finally, by Chebyshev's inequality we have

$$\Pr\left(\sum_j |d - d_j|^2 > 2dn\right) \leqslant \frac{21}{n},$$

which proves the second part of the lemma.

We now consider the sum of the variables $|\sqrt{d} - \sqrt{d_j}|^2$. We have

$$\begin{aligned}\sum_{j\in V} |\sqrt{d} - \sqrt{d_j}|^2 &= \sum_{i\in V} d + \sum_{i\in V} d_i - 2\sqrt{d}\cdot\sum_{j\in V}\sqrt{d_j} \\ &\leqslant 2dn + \sum_{i\in V} d_i - 2\sqrt{d}\cdot\sum_{j\in V}\sqrt{d_j}. \qquad (47)\end{aligned}$$

From the Chernoff bound (Lemma 76) we have that for some positive constant c_{cb} it holds w.h.p.

$$\sum_{j\in V} d_j = \sum_{\substack{u,v\in V\\ u\neq v}} 2e_{u,v} + \sum_{u\in V} e_{u,v} \leqslant 2dn + c_{\mathrm{cb}}\sqrt{dn\log n} \leqslant 4dn + n,$$

where we are using the hypothesis $d = o(n/\log n)$. We now prove that

$$\sum_{j\in V}\sqrt{d_j} \geqslant 2n\sqrt{d} - \frac{n}{\sqrt{d}},$$

which together with (47) implies that

$$\sum_{j\in V} |\sqrt{d} - \sqrt{d_j}|^2 \leqslant 4n,$$

concluding the proof of the lemma.

Observe that if $x \geqslant 0$, we have

$$\sqrt{x} \geqslant 1 + \frac{x-1}{2} - \frac{(x-1)^2}{2},$$

so that if X is a non-negative random variable of expectation 1 we have[8]

$$\mathbb{E}[\sqrt{X}] \geqslant 1 - \frac{\mathrm{Var}\,(X)}{2}.$$

By applying the above inequality to d_j/d we get

$$\mathbb{E}\left[\sqrt{\frac{d_j}{d}}\right] \geqslant 1 - \frac{\mathrm{Var}\left(\frac{d_j}{d}\right)}{2} = 1 - \frac{\mathrm{Var}\,(d_j)}{2d^2} \geqslant 1 - \frac{1}{2d},$$

and

$$\mathbb{E}[\sqrt{d_j}] \geqslant \sqrt{d} - \frac{1}{2\sqrt{d}}. \qquad (48)$$

[8]This argument is due to Ori Gurel-Gurevich (see **[GG]**).

We show that $\sum_{j\in V}\sqrt{d_j}$ is concentrated around its expectation by using Chebyshev's inequality[9]. In order to do that, we bound their covariance as

$$\mathbb{E}[\sqrt{d_i d_j}] - \mathbb{E}[\sqrt{d_i}]\mathbb{E}[\sqrt{d_j}] \leqslant \frac{8d^2}{n}.$$

By the law of total probability

$$\mathbb{E}[\sqrt{d_i}] = \Pr(e_{i,j})\mathbb{E}[\sqrt{d_i^{(j)}+1}] + (1-\Pr(e_{i,j}))\mathbb{E}[\sqrt{d_i^{(j)}}],$$

and

$$\begin{aligned}\mathbb{E}[\sqrt{d_j d_i}] = \Pr(e_{i,j})&\mathbb{E}[\sqrt{d_i^{(j)}+1}]\mathbb{E}[\sqrt{d_j^{(i)}+1}]\\ &+ (1-\Pr(e_{i,j}))\mathbb{E}[\sqrt{d_j^{(i)}}]\mathbb{E}[\sqrt{d_i^{(j)}}],\end{aligned}$$

which imply that

$$\begin{aligned}&\mathbb{E}[\sqrt{d_i d_j}] - \mathbb{E}[\sqrt{d_i}]\mathbb{E}[\sqrt{d_j}]\\
&= \Pr(e_{i,j})\mathbb{E}[\sqrt{d_i^{(j)}+1}]\mathbb{E}[\sqrt{d_j^{(i)}+1}]\\
&\quad + (1-\Pr(e_{i,j}))\mathbb{E}[\sqrt{d_j^{(i)}}]\mathbb{E}[\sqrt{d_i^{(j)}}]\\
&\quad - \Pr(e_{i,j})^2\mathbb{E}[\sqrt{d_j^{(i)}+1}]\mathbb{E}[\sqrt{d_i^{(j)}+1}]\\
&\quad - \Pr(e_{i,j})(1-\Pr(e_{i,j}))\mathbb{E}[\sqrt{d_j^{(i)}}]\mathbb{E}[\sqrt{d_i^{(j)}+1}]\\
&\quad - \Pr(e_{i,j})(1-\Pr(e_{i,j}))\mathbb{E}[\sqrt{d_j^{(i)}+1}]\mathbb{E}[\sqrt{d_i^{(j)}}]\\
&\quad - (1-\Pr(e_{i,j}))^2\mathbb{E}[\sqrt{d_j^{(i)}}]\mathbb{E}[\sqrt{d_i^{(j)}}]\\
&= p(1-p)\Big(\mathbb{E}[\sqrt{d_i^{(j)}+1}]\mathbb{E}[\sqrt{d_j^{(i)}+1}]\\
&\quad + \mathbb{E}[\sqrt{d_j^{(i)}}]\mathbb{E}[\sqrt{d_i^{(j)}}] + \mathbb{E}[\sqrt{d_j^{(i)}}]\mathbb{E}[\sqrt{d_i^{(j)}+1}]\\
&\quad + \mathbb{E}[\sqrt{d_j^{(i)}+1}]\mathbb{E}[\sqrt{d_i^{(j)}}]\Big) \leqslant \frac{8d^2}{n},\end{aligned} \tag{49}$$

where in the last inequality we used that by the Chernoff bound (Lemma 76) it holds w.h.p. $\mathbb{E}[\sqrt{d_i^{(j)}}] < \sqrt{2d}$, and that $p(1-p) < p < d/n$. From (49) it then follows that

$$\operatorname{Var}\left(\sum_{j\in V}\sqrt{d_j}\right) \leqslant 2nd + 32d^2n < \frac{n^2}{dn^{c_{\text{tight}}}}. \tag{50}$$

[9] A stronger bound which doesn't require the hypothesis $d \leqslant n^{1/3-c_{\text{tight}}}$ may be obtained with some concentration techniques compatible with the stochastic dependence among the $\sqrt{d_j}$s.

Finally, by combining (50) and (48) with Chebyshev's inequality we get

$$\begin{aligned}&\Pr(\sum_{j\in V}\sqrt{d_j} < 2n\sqrt{d} - \frac{n}{\sqrt{d}})\\ &\leqslant \Pr(\Big|\sum_{j\in V}\sqrt{d_j} - \mathbf{E}\Big[\sum_{j\in V}\sqrt{d_j}\Big]\Big| > \frac{n}{\sqrt{d}}) \leqslant \frac{1}{n^{c_{\mathrm{tight}}}}.\end{aligned}$$

gg

CHAPTER 5

3-Majority Dynamics

In this chapter we prove the results presented in Section 2.2. We consider two fundamental distributed consensus problems, in the setting in which each node in a complete communication network of size n initially holds an *opinion* (*color*), which is chosen arbitrarily from a finite set Σ. In the consensus problem the system must converge toward a consensus state in which all, or almost all nodes, hold the same opinion. Moreover, this opinion should be *valid*, i.e., it should be one among those initially present in the system. We further require this condition to be met even in the presence of a malicious adversary who can modify the opinions of a bounded subset of nodes, adaptively chosen in every round. In the more restrictive *plurality* consensus problem, the goal is having the process to converge to the *stable* configuration in which all nodes support the initial plurality.

In order to elegantly solve these problems, we study the *3-Majority dynamics*: At every round, every node pulls the opinion from three random neighbors and sets her new opinion to the majority one (ties are broken arbitrarily).

Let k be the number of valid opinions. As for the consensus problem, we show that, if $k \leqslant n^{\alpha}$, where α is a suitable positive constant, the 3-Majority dynamics converges in time polynomial in k and $\log n$, w.h.p., even in the presence of an adversary who can affect up to $o(\sqrt{n})$ nodes at each round. As for the plurality consensus problem, if the initial opinion configuration exhibits a sufficiently large *bias* s towards a fixed plurality opinion (that is, the number of nodes supporting the plurality opinion exceeds the number of nodes supporting any other opinion by s additional nodes), we prove that the 3-Majority dynamics converges in time $\mathcal{O}(\min\{k, (n/\log n)^{1/3}\}\log n)$, w.h.p. provided that $s \geqslant c\sqrt{\min\{2k, (n/\log n)^{1/3}\}\, n \log n}$. We then prove that our upper bound above is tight as long as $k \leqslant (n/\log n)^{1/4}$.

Finally, a natural question is whether looking at more (than three) random neighbors can significantly speed up the process. We provide a negative answer to this question: In particular, we show that samples of polylogarithmic size can speed up the process by a polylogarithmic factor only.

5.0.1. The majority roadmap

Section 5.1 formalizes the basic concepts and gives some preliminary results. Section 5.2 is devoted to the proofs of the upper bounds on the

convergence time of the 3-Majority dynamics. In Section 5.3, the lower bounds for the studied dynamics are described.

5.1. The 3-Majority Dynamics for Plurality Consensus

A *(k-opinion) configuration* (*k-cd* for short) is any k-tuple $\mathbf{c} = (c_1, \ldots, c_k)$ such that c_js are non negative integers and $\sum_{j=1,\ldots,k} c_j = n$. In what follows, we always assume w.l.o.g. $c_1 \geqslant c_2 \geqslant \cdots \geqslant c_k$. So c_1 is the *plurality opinion* and $s(\mathbf{c}) = c_1 - c_2$ is the *bias* of $\mathbf{c}$.

The 3-Majority dynamics works as follows:

> *At every round, every node samples three nodes (including herself and with repetitions) independently and uniformly at random and reset her opinion according to the majority of the opinions she sees. If she sees three different opinions, she chooses the first one.*

Clearly, in the case of three different opinions, choosing the second or the third one would not make any difference. The same holds even if the choice would be uniformly at random among the three opinions.

For any round t and for any $j \in [k]$, let $C_j^{(t)}$ be the r.v. counting the number of nodes with opinion j at round t and let $\mathbf{C}^{(t)} = (C_1^{(t)}, \ldots, C_k^{(t)})$ denote the random variable indicating the k-cd at time t of the execution of the 3-Majority dynamics.

For every $j \in [k]$ let $\mu_j(\mathbf{c})$ be the expected number of nodes with opinion j at the next round when the current k-cd is $\mathbf{c}$, i.e.

$$\mu_j(\mathbf{c}) = \mathbb{E}\left[C_j^{(t+1)} \,\middle|\, \mathbf{C}^{(t)} = \mathbf{c} \right].$$

To simplify the notation, in all the technical proofs we write μ_j and s instead of $\mu_j(\mathbf{c})$ and $s(\mathbf{c})$ when the dependence on configuration $\mathbf{c}$ is clear from the context.

LEMMA 17 (Next Expected Configuration). *For any k-cd* $\mathbf{c}$ *and for every opinion* $j \in [k]$*, it holds that*

$$\mu_j(\mathbf{c}) = c_j\left(1 + \frac{c_j}{n} - \sum_{h\in[k]} \frac{c_h^2}{n^2}\right). \tag{51}$$

PROOF. According to the 3-Majority dynamics, a node i gets opinion j if it chooses three times opinion j, or if it chooses two times j and one time a different opinion, or if it chooses the first time opinion j and then, the second and third time, two different distinct opinions. Hence if we name $X_{i,j}^{(t)}$ the indicator random variable of the event "Node i gets opinion j at time t", we

have that

$$\begin{aligned}
&P\left(X_{i,j}^{(t+1)} = 1 \,|\, \mathbf{C}^{(t)} = \mathbf{c}\right) \\
&= \left(\frac{c_j}{n}\right)^3 + 3\left(\frac{c_j}{n}\right)^2\left(\frac{n-c_j}{n}\right) \\
&\quad + \left(\frac{c_j}{n}\right)\left(1 - \left(\frac{\sum_{h=1}^k c_h^2}{n^2} + 2\left(\frac{c_j}{n}\right)\left(\frac{n-c_j}{n}\right)\right)\right) \\
&= \left(\frac{c_j}{n^3}\right)\left(n^2 + c_j n - \sum_{h=1}^k c_h^2\right).
\end{aligned}$$

gg

LEMMA 18 (Next expected bias). *For any k-cd $\mathbf{c}$ and for every opinion $j \in [k]$ with $j \neq 1$, it holds that*

$$\mu_1(\mathbf{c}) - \mu_j(\mathbf{c}) \geqslant s(\mathbf{c})\left(1 + \frac{c_1}{n}\left(1 - \frac{c_1}{n}\right)\right). \tag{52}$$

PROOF. Observe that, when we assume $c_1 \geqslant c_2 \geqslant \cdots \geqslant c_k$, we can give the following upper bound on the sum of squares in Lemma 17

$$\sum_{h\in[k]} c_h^2 = c_1^2 + \sum_{h=2}^k c_h^2 \leqslant c_1^2 + c_2 \sum_{h=2}^k c_h = c_1^2 + c_2(n - c_1). \tag{53}$$

From Lemma 17 it thus follows that, for any $j \neq 1$,

$$\begin{aligned}
\mu_1 - \mu_j \geqslant \mu_1 - \mu_2 &= (c_1 - c_2) + \frac{(c_1^2 - c_2^2)}{n} - \frac{c_1 - c_2}{n^2}\sum_{h\in k} c_h^2 \\
&= s \cdot \left(1 + \frac{c_1 + c_2}{n} - \frac{1}{n^2}\sum_{h\in k} c_h^2\right) \\
&\geqslant s \cdot \left(1 + \frac{c_1 + c_2}{n} - \frac{c_1^2 + n c_2}{n^2}\right) \\
&= s \cdot \left(1 + \frac{c_1}{n}\left(1 - \frac{c_1}{n}\right)\right),
\end{aligned}$$

where in the inequality we used (53) and the fact that $c_1 - c_2 \geqslant 0$. gg

5.2. Upper Bounds for 3-Majority Dynamics

In this section, we provide the following upper bound on the convergence time of the 3-Majority dynamics which clarifies the roles played by the plurality opinion and by the initial bias.

THEOREM 5 (General Upper Bound for 3-Majority). *Let λ be any value such that $\lambda < \sqrt[3]{n}$ and let $\mathbf{c}$ be any initial k-cd, with $c_1 \geqslant n/\lambda$ and*

$$s(\mathbf{c}) \geqslant 72\sqrt{2\lambda\, n \log n}.$$

Then the 3*-Majority dynamics converges to the plurality opinion in* $\mathcal{O}(\lambda \log n)$ *time w.h.p.*

The next three corollaries of Theorem 5 address three relevant special cases. Corollary 3 is obtained by setting $\lambda = \min\left\{2k,\ \sqrt[3]{n/\log n}\right\}$ and it provides a bound which does not assume any condition on c_m.

COROLLARY 3 (Upper Bound with Bias). *Let* $\mathbf{c}$ *be any initial* k*-color configuration with*

$$s(\mathbf{c}) \geqslant 72\sqrt{2\min\left\{2k,\ \sqrt[3]{\frac{n}{\log n}}\right\} n\log n}.$$

Then, the 3*-Majority dynamics converges to the plurality opinion in* $\mathcal{O}(\min\{2k, \sqrt[3]{n/\log n}\}\log n)$ *time w.h.p.*

Corollaries 4 and 7 are obtained by setting $\lambda = \operatorname{poly}\log(n)$ and $\lambda = \Theta(1)$, respectively. They provide sufficient conditions for a polylogarithmic convergence time.

COROLLARY 4 (Polylogarithmic Upper Bound for 3-Majority). *Let* $\mathbf{c}$ *be any initial* k*-cd with* $c_1 \geqslant n/\log^{\ell} n$ *and*

$$s(\mathbf{c}) \geqslant 72\sqrt{2n\log^{\ell+1} n}.$$

Then, the 3*-Majority dynamics converges to the plurality opinion in* $\mathcal{O}(\log^{\ell+1} n)$ *time w.h.p.*

COROLLARY 7 (Logarithmic Upper Bound for 3-Majority). *Let* $\mathbf{c}$ *be any* k*-cd with* $c_1 \geqslant n/\beta$ *and* $s(\mathbf{c}) \geqslant 72\sqrt{2\beta n\log n}$*, for some constant* $\beta \geqslant 1$*. Then, the* 3*-Majority dynamics converges to the plurality opinion in* $\mathcal{O}(\log n)$ *rounds, w.h.p.*

In order to prove Theorem 5, we need the following three technical lemmas that essentially characterize three different phases of the process analysis. Each of them concerns a different range assumed by the plurality c_1. The first lemma considers configurations in which c_1 is less than a suitable constant fraction of n: in this case, it shows that the bias between the plurality size and the size of any other opinion increases by a factor $1 + \Omega(c_1/n) = 1 + \Omega(1/\lambda)$.

LEMMA 19 (From Plurality to Majority). *Let* $\mathbf{c}$ *be any* k*-cd with* $n/\lambda \leqslant c_1 \leqslant 2n/3$ *and* $s(\mathbf{c}) \geqslant 72\sqrt{2\lambda n\log n}$ *where* $\lambda < \sqrt[3]{n}$*. and* α *is a sufficiently large constant. Then, for any other opinion* $j \neq 1$ *it holds that*

$$\Pr\left(C_1^{(t+1)} - C_j^{(t+1)} \geqslant s(\mathbf{c})\left(1 + \frac{c_1}{4n}\right) \,\middle|\, \mathbf{C}^{(t)} = \mathbf{c}\right) \geqslant 1 - \frac{1}{n^3}.$$

PROOF. Conditional on any configuration $\mathbf{C}^{(t)} = \mathbf{c}$, from the Chernoff bounds (Lemma 76, in particular (188) with $\delta = 3\sqrt{\log n/\mu}$ if $\mu > \log n$,

(189) with $\delta = 4\log n/\mu$ otherwise), it follows that w.h.p.

$$\begin{aligned} C_j^{(t+1)} &\leqslant \max\left\{\mu_j + 3\sqrt{\mu_j \log n}, 5\log n\right\}, \\ C_1^{(t+1)} &\geqslant \mu_1 - 3\sqrt{\mu_1 \log n}. \end{aligned}$$

Thus, if $\mu_j + 3\sqrt{\mu_j \log n} \geqslant 5\log n$, then it holds w.h.p.

$$\begin{aligned} C_1^{(t+1)} - C_j^{(t+1)} &\geqslant \mu_1 - \mu_j - 3\sqrt{\mu_1 \log n} - 3\sqrt{\mu_j \log n} \\ &\geqslant \mu_1 - \mu_j - 2\alpha\sqrt{\mu_1 \log n}, \end{aligned} \tag{54}$$

where we used that by the union bound $\Pr(A \cap B) \geqslant 1 - \Pr(A^C) - \Pr(B^C)$. Otherwise, if $\mu_j + 3\sqrt{\mu_j \log n} < 5\log n$, then it holds w.h.p.

$$\begin{aligned} C_1^{(t+1)} - C_j^{(t+1)} &\geqslant \mu_1 - 3\sqrt{\mu_1 \log n} - 5\log n \\ &\geqslant \mu_1 - \mu_j - 6\sqrt{\mu_1 \log n}, \end{aligned} \tag{55}$$

where in the last inequality we used that $\mu_1 \geqslant c_1 \geqslant n/\lambda \geqslant n^{\frac{2}{3}}$.

From Lemma 18 and the hypothesis $c_1 \leqslant 2n/3$ we get that

$$\mu_1 - \mu_j \geqslant (c_1 - c_j)\left(1 + \frac{c_1}{3n}\right),$$

and from (51) we also have that $\mu_1 \leqslant 2c_1$. Thus, in (54) and (55) we get

$$\begin{aligned} \mu_1 - \mu_j - 6\sqrt{\mu_1 \log n} &\geqslant (c_1 - c_j)\left(1 + \frac{c_1}{3n}\right) - 6\sqrt{2c_1 \log n} \\ &\geqslant (c_1 - c_j)\left(1 + \frac{c_1}{3n} - 6\frac{\sqrt{2c_1 \log n}}{(c_1 - c_j)}\right) \\ &\overset{(a)}{\geqslant} (c_1 - c_j)\left(1 + \frac{c_1}{3n} - \frac{1}{12}\sqrt{\frac{c_1}{\lambda n}}\right) \\ &\geqslant (c_1 - c_j)\left(1 + \frac{c_1}{3n}\left(1 - \frac{1}{4}\sqrt{\frac{n}{c_1\lambda}}\right)\right) \\ &\overset{(b)}{\geqslant} (c_1 - c_j)\left(1 + \frac{c_1}{4n}\right), \end{aligned}$$

where in (a) we used that $c_1 - c_j \geqslant s \geqslant 72\sqrt{2\lambda n \log n}$ and in (b) we used that $c_1 \geqslant n/\lambda$, concluding the proof. ▣

Once c_1 becomes larger than $2n/3$ the negative occurrence of c_1 in (52) does not allow to directly show a drift towards plurality. We thus consider another useful "drift" of the process: The sum of all the other opinion sizes decreases exponentially, w.h.p., as long as this sum is enough large to apply concentration bounds. This result is formalized in the next lemma.

LEMMA 20 (From majority to almost all). *Let* $\mathbf{c}$ *be any* k*-cd with* $2n/3 \leqslant c_1 \leqslant n - \omega(\log n)$*. Then, it holds that*

$$\Pr\left(\sum_{i\neq 1} C_i^{(t+1)} \leqslant \frac{8}{9}\sum_{i\neq 1} c_i \,\middle|\, \mathbf{C}^{(t)} = \mathbf{c}\right) \geqslant 1 - \frac{1}{n^3}.$$

PROOF. Let us define $\mu_{-1} = \sum_{i\neq 1}\mu_i$. From (51) we have

$$\begin{aligned}
\frac{\mu_{-1}}{n} &= \sum_{i\neq 1}\frac{c_i}{n}\left(1 + \frac{c_i}{n} - \sum_j \left(\frac{c_j}{n}\right)^2\right)\\
&= 1 - \frac{c_1}{n} + \sum_{i\neq 1}\left(\frac{c_i}{n}\right)^2 - \left(1 - \frac{c_1}{n}\right)\sum_j\left(\frac{c_j}{n}\right)^2\\
&= 1 - \frac{c_1}{n} - \left(\frac{c_1}{n}\right)^2 + \frac{c_1}{n}\sum_j\left(\frac{c_j}{n}\right)^2\\
&\overset{(a)}{\leqslant} 1 - \frac{c_1}{n} - \left(\frac{c_1}{n}\right)^2 + \frac{c_1}{n}\left(\left(\frac{c_1}{n}\right)^2 + \frac{c_2}{n}\left(1 - \frac{c_1}{n}\right)\right)\\
&= \left(1 - \frac{c_1}{n}\right)\left(1 - \left(\frac{c_1}{n}\right)^2 + \frac{c_1}{n}\frac{c_2}{n}\right)\\
&= \left(1 - \frac{c_1}{n}\right)\left(1 - \frac{c_1}{n}\left(\frac{c_1}{n} - \frac{c_2}{n}\right)\right), \qquad (56)
\end{aligned}$$

where in (a) we used (53). Using the hypothesis $c_1/n \geqslant 2/3$ (hence $c_2/n \leqslant 1/3$), from (56) we obtain the last expression become

$$\left(1 - \frac{c_1}{n}\right)\left(1 - \frac{c_1}{n}\left(\frac{c_1}{n} - \frac{c_2}{n}\right)\right) \leqslant \left(1 - \frac{c_1}{n}\right)\left(1 - \frac{c_1}{3n}\right) \leqslant \frac{7}{9}\sum_{i\neq 1}\frac{c_i}{n}. \qquad (57)$$

Now observe that, from the Chernoff bound (Lemma 76), as long as $\mu_{-1} \in \omega(\log n)$, it holds w.h.p.

$$\begin{aligned}
\sum_{i\neq 1} C_i^{(t+1)} &\leqslant \mu_{-1} + \sqrt{\mu_{-1}\log n} \qquad (58)\\
&= \mu_{-1}\left(1 + \sqrt{\frac{\log n}{\mu_{-1}}}\right) = \mu_{-1}(1 + o(1)).
\end{aligned}$$

Thus, by replacing (57) in (58), we get that it holds w.h.p.

$$\sum_{i\neq 1} C_i^{(t+1)} \leqslant \mu_{-1}(1 + o(1)) \leqslant \frac{8}{9}\sum_{i\neq 1} c_i,$$

concluding the proof. gg

Finally, when the sum of all the minority opinions is not larger than a polylogarithmic function, the probability that they all disappear in one round is high. This is shown in the next lemma.

LEMMA 21 (The last step). *Let $\alpha > 0$ and let $\mathbf{c}$ be any k-cd with $c_1 \geqslant n - \log^\alpha n$. Then, it holds that*

$$\Pr\left(\sum_{i\neq 1} C_i^{(t+1)} = 0 \,|\, \mathbf{C}^{(t)} = \mathbf{c}\right) \geqslant 1 - \frac{3\log^{2\alpha} n}{n}. \tag{59}$$

PROOF. As in the previous proof let we name $\mu_{-1} = \sum_{i\neq 1}\mu_i$. Note that $c_1 \geqslant n - \log^\alpha n$ implies $\sum_{i\neq 1} c_i \leqslant \log^\alpha n$. Thus, from (51) we have

$$\begin{aligned}
\mu_{-1} &= \sum_{i\neq 1} c_i \left(1 + \frac{c_i}{n} - \sum_j \left(\frac{c_j}{n}\right)^2\right) \\
&\leqslant \sum_{i\neq 1} c_i \left(1 + \frac{c_i}{n} - \left(\frac{c_1}{n}\right)^2\right) \\
&= \sum_{i\neq 1} c_i \left(1 + \frac{c_i}{n} - \left(1 - \sum_{j\neq 1}\frac{c_j}{n}\right)^2\right) \\
&\leqslant \sum_{i\neq 1} c_i \left(\frac{c_i}{n} + 2\sum_{j\neq 1}\frac{c_j}{n}\right) \\
&\leqslant \sum_{i\neq 1} c_i \left(\frac{3\log^\alpha n}{n}\right) = \frac{3\log^{2\alpha} n}{n}.
\end{aligned}$$

Finally, (59) follows from Markov's inequality on the event "$\sum_{i\neq 1} C_i^{(t+1)} \geqslant 1$" and, since $\sum_{i\neq 1} C_i^{(t+1)}$ is a non-negative integer-valued r.v., this is equivalent as "$\sum_{i\neq 1} C_i^{(t+1)} > 0$". ◻◻

PROOF OF THEOREM 5. From Lemma 19 it follows that, as long as the number of nodes with the plurality opinion c_1 is smaller than a constant fraction of n, the bias between c_1 and c_2 increases by a factor $(1 + \frac{1}{4\lambda})$, w.h.p.

From Lemma 20 it follows that, when the plurality opinion reaches a suitable constant fraction of n, then the number of nodes with non-plurality opinions decreases at exponential rate, w.h.p.

Finally, in Lemma 21 we consider separately the last round of the protocol, where all opinions but the plurality one disappear, w.h.p. ◻◻

5.2.1. Plurality consensus with adversary

In this section we show that the 3-Majority dynamics is robust against Byzantine adversaries. Let $F \leqslant n$, we consider an *F-bounded dynamic adversary* that, at every round, can change the opinion of up to F nodes with the goal of preventing the system to converge to the plurality opinion.

Clearly, reaching complete plurality consensus is not possible in this framework. In presence of an F-bounded dynamic adversary we thus consider the *M-plurality consensus*, in which all but M nodes have to agree on the plurality opinion.

Notice that it is not possible to reach M-plurality consensus against an F-bounded dynamic adversary if $F > M$. Our previous analysis of the 3-Majority dynamics can be easily adapted to show that it achieves $o(s/\lambda)$-plurality consensus against any F-bounded adversary for $F = o(s/\lambda)$, where s is the initial bias and $\lambda < \sqrt[3]{n}$.

COROLLARY 5 (Upper Bound with Adversary). *Let λ be any value such that $\lambda < \sqrt[3]{n}$ and let $\mathbf{c}$ be any initial configuration, with $c_1 \geqslant n/\lambda$ and*

$$s(\mathbf{c}) \geqslant 24\sqrt{2\lambda\, n \log n}.$$

The 3-Majority dynamics achieves $\mathcal{O}(s(\mathbf{c})/\lambda)$-plurality consensus against any F-bounded adversary with $F = o(s(\mathbf{c})/\lambda)$, and the convergence time is $\mathcal{O}(\lambda \log n)$ w.h.p.

PROOF. In order to formalize the analysis of the process with an adversary, we split each round in two consecutive steps: In the first step nodes apply the updating rule of the 3-Majority dynamics while, in the second step, the adversary can change the opinion of up to F arbitrary nodes. Hence, if the configuration of the system at some round t is $\mathbf{C}^{(t)} = \hat{\mathbf{c}}$, we name $\mathbf{H}^{(t+1)}$ the random variable indicating the configuration after the first step of round $t+1$ and $\mathbf{C}^{(t+1)}$ the configuration after the second step of round $t+1$, i.e.

$$\mathbf{C}^{(t)} = \hat{\mathbf{c}} \overset{\mathcal{R}andom}{\longrightarrow} \mathbf{H}^{(t+1)} \overset{\mathcal{A}dversary}{\longrightarrow} \mathbf{C}^{(t+1)}.$$

Notice that $\mathbf{C}^{(t+1)}$ is a function of $\mathbf{H}^{(t+1)}$ arbitrarily determined by the adversary within its constraints.

If the configuration at some round t is $\mathbf{C}^{(t)} = \hat{\mathbf{c}}$, with $\hat{c}_1 \leqslant 2n/3$, then from Lemma 19 it follows that w.h.p.

$$H_1^{(t+1)} - H_j^{(t+1)} \geqslant s(\hat{\mathbf{c}}) + \frac{s(\hat{\mathbf{c}})}{4\lambda}.$$

The bias after the adversarial step is thus w.h.p.

$$C_1^{(t+1)} - C_j^{(t+1)} \geqslant s(\hat{\mathbf{c}}) + s(\hat{\mathbf{c}})/(4\lambda) - F.$$

Since by hypothesis $F = \mathcal{O}(s(\mathbf{c})/\lambda)$, as long as the bias $s(\hat{\mathbf{c}})$ of the current configuration is at least as large as the bias $s(\mathbf{c})$ of the initial configuration, we have that w.h.p.

$$C_1^{(t+1)} - C_j^{(t+1)} \geqslant s(\hat{\mathbf{c}}) + \frac{s(\hat{\mathbf{c}})}{4\lambda} - F \geqslant s(\hat{\mathbf{c}}) + \frac{s(\hat{\mathbf{c}})}{5\lambda}. \tag{60}$$

Notice that the requirement $s(\hat{\mathbf{c}}) \geqslant s(\mathbf{c})$ trivially holds in the initial configuration, when $\hat{\mathbf{c}} = \mathbf{c}$, and from 60 by induction it holds in all the following rounds, w.h.p.

(60) guarantees that, as long as the plurality opinion is supported by at most $2n/3$ nodes (see hypothesis of Lemma 19) the bias increases by a factor $1 + \Theta(1/\lambda)$ at each round, w.h.p., even in the presence of the adversary. Hence, after $\mathcal{O}(\lambda \log n)$ rounds the plurality opinion is supported by at least $2n/3$ nodes, w.h.p.

When the system reaches, at some round t, a configuration $\mathbf{C}^{(t)} = \hat{\mathbf{c}}$ such that the plurality opinion is supported by $2n/3 \leqslant \hat{c}_1 \leqslant n - \omega(\log n)$ nodes, then Lemma 20 guarantees that the total number of nodes supporting the other opinions in configuration $\mathbf{H}^{(t+1)}$ after the step of 3-Majority dynamics of the next round is w.h.p.

$$\sum_{i \neq 1} H_i^{(t+1)} \leqslant \frac{8}{9} \sum_{i \neq 1} \hat{c}_1 \,.$$

Hence, as long as $\sum_{i \neq 1} \hat{c}_1 = \Omega(s(\mathbf{c})/\lambda)$, the total number of nodes supporting the other opinions in configuration $\mathbf{C}^{(t+1)}$ (after the adversarial step of the next round) is w.h.p.

$$\sum_{i \neq 1} C_i^{(t+1)} \leqslant \frac{8}{9} \sum_{i \neq 1} \hat{c}_1 + F \leqslant \frac{9}{10} \sum_{i \neq 1} \hat{c}_1 \,. \tag{61}$$

Thus, when the plurality opinion reaches $2n/3$ nodes, after further $\mathcal{O}(\log n)$ rounds all but $o(s(\mathbf{c})/\lambda)$ nodes support the plurality opinion, w.h.p. Notice that (61) also guarantees that, once we reached M-plurality consensus, the system takes on only configurations that satisfy M-plurality consensus, w.h.p. ▣

5.3. Lower Bounds for 3-Majority Dynamics

This section is organized in three subsections:

- In Section 5.3.1, we prove a lower bound on the convergence time of the 3-Majority dynamics;
- In Section 5.3.2, we show that the 3-Majority dynamics is essentially the only 3-input dynamics that converges to plurality consensus;
- In Section 5.3.3, we provide a lower bound on the convergence time of the h-plurality dynamics for $h > 3$.

5.3.1. Lower bound for 3-Majority dynamics

In this section we show that if the 3-Majority dynamics starts from a sufficiently balanced configuration (i.e., at the beginning there are $n/k \pm o(n/k)$ nodes of every opinion) then it takes $\Omega(k \log n)$ rounds, w.h.p., to reach one of the absorbing configurations where all nodes have the same opinion. In what follows, all events and random variables thus concern the Markov process yielded by the 3-Majority dynamics.

In the next lemma we show that if there are at most $n/k + b$ nodes of a specific opinion, where b is smaller than n/k, then at the next round there are at most $n/k + (1 + 3/k)b$ nodes of that opinion, w.h.p.

LEMMA 22. *Let the number of opinions k be such that $k \leqslant (n/\log n)^{1/4}$, let b be any number with $k\sqrt{n\log n} \leqslant b \leqslant n/k$, and let $\mathbf{c} = (c_1, \ldots, c_k)$ be a configuration. If $c_j = n/k + a$ for some opinion $j \in [k]$ and for some $a \leqslant b$, then the number of nodes with opinion j at the next round are at most $n/k + (1 + 3/k)b$, w.h.p.; more precisely, for any $a \leqslant b$ and for any configuration $\mathbf{c}$ such that $c_j = n/k + a$ it holds that*

$$\Pr\left(C_j^{(t+1)} \geqslant \frac{n}{k} + \left(1 + \frac{3}{k}\right) b \;\middle|\; \mathbf{C}^{(t)} = \mathbf{c}\right) \leqslant \frac{1}{n^2}.$$

PROOF. For any configuration $\mathbf{c} = (c_1, \ldots, c_k)$ with $\sum_{j=1}^k c_j = n$ and any opinion $j \in [k]$, the expected value of the number of nodes having opinion j at round $t+1$ conditional on $\left\{\mathbf{C}^{(t)} = \mathbf{c}\right\}$ is (see Lemma 17)

$$\mathbb{E}\left[C_j^{(t+1)} \mid \mathbf{C}^{(t)} = \mathbf{c}\right] = c_j \left(1 + \frac{c_j}{n} - \frac{1}{n^2}\sum_{j=1}^k c_j^2\right).$$

Observe that, since $\sum_{j=1}^k c_j = n$, from Jensen's inequality[1] it follows that

$$\sum_{j=1}^k \frac{c_j^2}{n^2} \geqslant \frac{1}{k}.$$

Hence, we can give an upper bound on the expectation of $C_j^{(t+1)}$ that depends only on c_j and not on the whole configuration $\mathbf{c}$ at round t, namely

$$\mathbb{E}\left[C_j^{(t+1)} \mid \mathbf{C}^{(t)}\right] \leqslant C_j^{(t)} \left(1 + \frac{C_j^{(t)}}{n} - \frac{1}{k}\right).$$

If we condition on the number of nodes of opinion j being $c_j = n/k + a$ in configuration $\mathbf{c}$, for some $a \leqslant b$, we get

$$\begin{aligned}
\mathbb{E}\left[C_j^{(t+1)} \mid \mathbf{C}^{(t)} = \mathbf{c}\right] &\leqslant \left(\frac{n}{k} + a\right)\left(1 + \frac{n/k + a}{n} - \frac{1}{k}\right) \\
&= \frac{n}{k} + \left(1 + \frac{1}{k}\right) a + \frac{a^2}{n} \\
&\leqslant \frac{n}{k} + \left(1 + \frac{1}{k}\right) b + \frac{b^2}{n} \leqslant \frac{n}{k} + \left(1 + \frac{2}{k}\right) b,
\end{aligned}$$

where in the last two inequalities we used that $a \leqslant b$ and $b \leqslant n/k$.[2] Since $C_j^{(t+1)}$ conditional on $\left\{\mathbf{C}^{(t)} = \mathbf{c}\right\}$ can be written as a sum of n independent

[1] Jensen's inequality states that given any convex function $\phi : \mathbb{R} \to \mathbb{R}$ and k real numbers $x_1, \ldots x_k \in \mathbb{R}$, it holds $\phi\left(\frac{1}{k}\sum_{i=1}^k x_i\right) \leqslant \frac{1}{k}\sum_{i=1}^k \phi(x_i)$.

[2] Notice that the inequality holds in particular for negative a as well

Bernoulli random variables, from the Chernoff bound (Lemma 76) we thus get that for every $a \leqslant b$ it holds that

$$\Pr\left(C_j^{(t+1)} \geqslant \frac{n}{k} + \left(1+\frac{3}{k}\right) b \;\middle|\; \mathbf{C}^{(t)} = \mathbf{c}\right) \leqslant e^{-2(b/k)^2/n} \leqslant \frac{1}{n^2},$$

where in the last inequality we used that $b \geqslant k\sqrt{n \log n}$. ⊠

Let us say that a configuration $\mathbf{c} = (c_1, \ldots, c_k) \in \{0, 1, \ldots, n\}^k$ with $\sum_{j=1}^k c_j = n$ is *monochromatic* if there is an $j \in [k]$ such that $c_j = n$. In the next theorem we show that if we start from a sufficiently *balanced* configuration, then the 3-Majority dynamics takes $\Omega(k \log n)$ rounds, w.h.p., to reach a monochromatic configuration.

THEOREM 6 (Lower Bound for 3-Majority). *Let*

$$\tau = \inf\{t \in \mathbb{N} : \mathbf{C}^{(t)} \textit{ is monochromatic}\}$$

be the random variable indicating the first round such that the system is in a monochromatic configuration. If the initial number of opinions is $k \leqslant (n/\log n)^{1/4}$ *and the initial configuration is* $\mathbf{c} = (c_1, \ldots, c_k)$ *with*

$$\max\{c_j : j = 1, \ldots, k\} \leqslant \frac{n}{k} + \left(\frac{n}{k}\right)^{1-\varepsilon}$$

for some $\varepsilon > 0$, *then* $\tau = \Omega(k \log n)$ *w.h.p.*

IDEA OF PROOF. For an opinion $j \in [k]$ let us denote the difference $C_j - n/k$ as the *positive imbalance*. In Lemma 22 we proved that, as long as the positive imbalance of an opinion is smaller than n/k, this difference increases by a factor smaller than $(1 + 3/k)$ at every round, w.h.p. Hence, if an opinion starts with a positive imbalance smaller than $(n/k)^{1-\varepsilon}$, for some $\varepsilon > 0$, then it takes $\Omega(k \log n)$ rounds to reach an imbalance of n/k, w.h.p. By union bounding on all the opinions, we can get the stated lower bound. ⊠

PROOF. Observe that for any round $T \leqslant c\, k \log n$, where c is a suitable positive constant, it holds that

$$(1 + 3/k)^T (n/k)^{1-\varepsilon} \leqslant \frac{n}{k}$$

Since in the initial configuration $\mathbf{c}$ for any opinion $j \in [k]$ we have that $c_j \leqslant n/k + (n/k)^{1-\varepsilon}$, for $T \leqslant c\, k \log n$ it holds that

$$\begin{aligned} &\Pr\left(C_j^{(T)} = n \;\middle|\; \mathbf{C}^{(0)} = \mathbf{c}\right) \\ &\leqslant \Pr\left(C_j^{(T)} \geqslant \frac{n}{k} + \left(1+\frac{3}{k}\right)^T \left(\frac{n}{k}\right)^{1-\varepsilon} \;\middle|\; \mathbf{C}^{(0)} = \mathbf{c}\right), \end{aligned} \tag{62}$$

Since $c_j \leqslant n/k + (n/k)^{1-\varepsilon}$, if we also have

$$C_j^{(T)} \geqslant \frac{n}{k} + \left(1+\frac{3}{k}\right)^T \left(\frac{n}{k}\right)^{1-\varepsilon},$$

then a round t with $0 \leqslant t \leqslant T-1$ must exist such that $C_j^{(t)} \leqslant n/k + b$ and

$$C_j^{(t+1)} \geqslant \frac{n}{k} + \left(1 + \frac{3}{k}\right) b$$

for some value b, with $k\sqrt{n \log n} \leqslant b \leqslant n/k$, thus

$$\Pr\left(C_j^{(T)} \geqslant \frac{n}{k} + \left(1+\frac{3}{k}\right)^T \left(\frac{n}{k}\right)^{1-\varepsilon} \,\middle|\, \mathbf{C}^{(0)} = \mathbf{c}\right) \tag{63}$$

$$\leqslant \Pr\left(\left(\exists t : 0 \leqslant t \leqslant T-1 \ \wedge\ C_j^{(t)} \leqslant \frac{n}{k} + b\right)\right.$$

$$\left. \wedge\ \left(C_j^{(t+1)} \geqslant \frac{n}{k} + \left(1+\frac{3}{k}\right) b\right) \,\middle|\, \mathbf{C}^{(0)} = \mathbf{c}\right) \tag{64}$$

$$\leqslant \sum_{t=0}^{T-1} \Pr\left(\left(C_j^{(t)} \leqslant \frac{n}{k} + b_t\right)\right.$$

$$\left. \wedge\ \left(C_j^{(t+1)} \geqslant \frac{n}{k} + \left(1+\frac{3}{k}\right) b_t\right) \,\middle|\, \mathbf{C}^{(0)} = \mathbf{c}\right) \tag{65}$$

where the inequality from (63) to (64) holds for some b with

$$k\sqrt{n \log n} \leqslant b \leqslant n/k,$$

and the inequality from (64) to (65) holds for some $b_0, \ldots, b_{T-1}$ with $k\sqrt{n \log n} \leqslant b_t \leqslant n/k$ for every $t = 0, \ldots, T-1$. Now observe that

(66)

$$\Pr\left(\left(C_j^{(t)} \leqslant \frac{n}{k} + b_t\right) \wedge \left(C_j^{(t+1)} \geqslant \frac{n}{k} + \left(1+\frac{3}{k}\right) b_t\right) \,\middle|\, \mathbf{C}^{(0)} = \mathbf{c}\right)$$

$$= \sum_{a \leqslant b_t} \Pr\left(\left(C_j^{(t)} = \frac{n}{k} + a\right) \wedge \left(C_j^{(t+1)} \geqslant \frac{n}{k} + \left(1+\frac{3}{k}\right) b_t\right) \,\middle|\, \mathbf{C}^{(0)} = \mathbf{c}\right)$$

$$= \sum_{a \leqslant b_t} \Pr\left(\left(C_j^{(t+1)} \geqslant \frac{n}{k} + \left(1+\frac{3}{k}\right) b_t\right) \,\middle|\, \left(C_j^{(t)} = \frac{n}{k} + a\right) \wedge \left(\mathbf{C}^{(0)} = \mathbf{c}\right)\right)$$

$$\cdot \Pr\left(C_j^{(t)} = \frac{n}{k} + a \,\middle|\, \mathbf{C}^{(0)} = \mathbf{c}\right)$$

$$\leqslant \frac{1}{n^2} \sum_{a \leqslant b_t} \Pr\left(C_j^{(t)} = \frac{n}{k} + a \,\middle|\, \mathbf{C}^{(0)} = \mathbf{c}\right) \leqslant \frac{1}{n^2},$$

where in the last line we used Lemma 22.

By combining (62), (65), and (66) we get that, for every opinion $j \in [k]$, if the initial number of nodes having opinion j is $c_j \leqslant n/k + (n/k)^{1-\varepsilon}$ at any round $T \leqslant c\,k \log n$ the probability that all nodes have opinion j is at

most T/n^2. The probability that $\mathbf{C}^{(T)}$ is monochromatic is thus at most $(kT)/n^2 \leqslant n^{-\alpha}$ for some positive constant α. ⊠

It may be worth noticing that what we actually prove in Theorem 6 is that $\Omega(k \log n)$ rounds are required in order to go from a configuration where the majority opinion has at most $n/k + (n/k)^{1-\varepsilon}$ nodes to a configuration where it has $2n/k$ opinions.

5.3.2. A negative result for 3-input dynamics

In order to prove that dynamics that differ from the majority ones do not solve plurality consensus, we first give some formal definitions of the dynamics we are considering.

Definition 10 (h-Input Dynamics). An *h-dynamics* is a synchronous protocol where at each round every node picks h random neighbors (including herself and with repetition) and updates her opinion according to some deterministic rule that depends only on the opinions it sees. Let $\mathcal{D}_h(k)$ be the class of h-dynamics and observe that a dynamics $\mathcal{P} \in \mathcal{D}_h$ can be specified by a function

$$f : [k]^h \to [k],$$

such that $f(x_1, \ldots, x_h) \in \{x_1, \ldots, x_h\}$, where $f(x_1, \ldots, x_h)$ is the opinion chosen by a node that sees the (ordered) sequence $(x_1, \ldots, x_h)$ of opinions.

In the class $\mathcal{D}_3(k)$, there is a subset $\mathcal{M}^3$ of equivalent protocols called 3-Majority dynamics having two key-properties described below: the clear-majority and the uniform one.

Definition 11 (Clear-Majority Property). Let $(x_1, x_2, x_3) \in [k]^3$ be a triple of opinions. We say that (x_1, x_2, x_3) has a *clear majority* if at least two of the three entries have the same value. A dynamics $\mathcal{P} \in \mathcal{D}_3(k)$ has the *clear-majority* property if whenever its f sees a clear majority it returns the majority opinion.

Given any 3-input dynamics function $f(x_1, x_2, x_3)$, for any triple of distinct opinions $r, g, b \in [k]$, let $\Pi(r, g, b)$ be the subset of permutations of the opinions r, g, b and define the following "counters":

$$\delta_r = |\{(z_1, z_2, z_3) \in \Pi(r, g, b),\ s.t.\ f(z_1, z_2, z_3) = r\}|,$$
$$\delta_g = |\{(z_1, z_2, z_3) \in \Pi(r, g, b),\ s.t.\ f(z_1, z_2, z_3) = g\}|,$$
$$\delta_b = |\{(z_1, z_2, z_3) \in \Pi(r, g, b),\ s.t.\ f(z_1, z_2, z_3) = b\}|.$$

Observe that for any 3-input dynamics it must hold $\delta_g + \delta_r + \delta_b = 6$.

Definition 12 (Uniform Property). A dynamics $\mathcal{P} \in \mathcal{D}_3(k)$ has the *uniform* property if, for any triple of distinct opinions $r, g, b \in [k]$, it holds that $\delta_r = \delta_g = \delta_b$ $(= 2)$.

Informally speaking, the clear-majority and the uniform properties provide a clean characterization of those dynamics that are good solvers for

plurality consensus. This fact is formalized in the next definitions and in the final theorem.

DEFINITION 13 (3-Input Majority-Boosting Dynamics). A protocol $\mathcal{P} \in \mathcal{D}_3(k)$ belongs to the class $\mathcal{M}^3 \subset \mathcal{D}_3(k)$ of *3-input majority-boosting dynamics* if its function $f(x_1, x_2, x_3)$ has the clear-majority and the uniform properties.

DEFINITION 14 ((s,ε)-Plurality Consensus Solver). We say that a protocol $\mathcal{P}$ is an (s,ε)-*solver* (for the plurality consensus problem) if for every initial s-biased configuration $\mathbf{c}$, when running $\mathcal{P}$, with probability at least $1-\varepsilon$ there is a round t by which all nodes get the plurality opinion of c.

Let us observe that, by definition of h-dynamics (see Definition 10), any monochromatic configuration is an absorbing state of the relative Markov process. Moreover, the smaller s and ε the better an (s,ε)-solver is; in other words, if a dynamics is an (s,ε)-solver then it is also an (s',ε')-solver for every $s' \geqslant s$ and $\varepsilon' \geqslant \varepsilon$. In Section 5.2, we showed that any dynamics

$$\mathcal{P} \in \mathcal{M}^3 \implies (\Theta(\sqrt{\min\{2k, (n/\log n)^{1/3}\} n \log n}), \Theta(1/n))\text{-solver} \in \mathcal{D}_3.$$

We can now state the main result of this section.

THEOREM 21 (Properties of Good Solvers). *Given a protocol $\mathcal{P}$, the following hold:*

- *(a) If $\mathcal{P}$ is an $(n/4, 1/4)$-solver in $\mathcal{D}_3$, then its f must have the clear-majority property.*
- *(b) A constant $\eta > 0$ exists such that, if $\mathcal{P}$ is an $(\eta \cdot n, 1/4)$-solver, then its f must have the uniform property.*

The above theorem also provides the clear reason why some dynamics can solve consensus but cannot solve plurality consensus in the non-binary case. A relevant example is the 3-Median dynamics studied in [**DGM$^+$11**]: it has the clear-majority property but not the uniform one.

For readability sake, we split the proof of the above theorem in two technical lemmas: in the first one, we show the first claim about clear majority while in the second lemma we show the second claim about the uniform property.

LEMMA 23 (clear majority). *If a protocol $\mathcal{P} \in \mathcal{D}_3$ is an $(n/4, 1/4)$-solver, then it chooses the majority opinion every time there is a triple with a clear majority.*

PROOF. For every triple of opinions $(x_1, x_2, x_3) \in [k]^3$ that has a clear majority, let us define $\delta(x_1, x_2, x_3)$ to be 1 if protocol $\mathcal{P}$ behaves like the majority protocol over triple (x_1, x_2, x_3) and 0 otherwise. Consider an initial configuration with only two opinions, say red (r) and blue (b), with c_r red

nodes and $c_b = n - c_r$ blue nodes. Let us define Δ_r and Δ_b as follows

$$\Delta_r = \delta(r,r,b) + \delta(r,b,r) + \delta(b,r,r),$$
$$\Delta_b = \delta(b,b,r) + \delta(b,r,b) + \delta(r,b,b).$$

We can write the probability that a node chooses opinion red as

$$p(r) = \left(\frac{c_r}{n}\right)^3 + \left(\frac{c_r}{n}\right)^2 \frac{c_b}{n} \cdot \Delta_r + \left(\frac{c_b}{n}\right)^2 \frac{c_r}{n} (3 - \Delta_b) \tag{67}$$
$$= \frac{c_r}{n^3} \left(c_r^2 + c_b \left(c_r \Delta_r - c_b \Delta_b\right) + 3c_b^2\right).$$

Observe that for a majority protocol we have that $\Delta_r = \Delta_b = 3$. In what follows we show that if this is not the case then there are configurations where the majority opinion does not increase in expectation. We distinguish two cases, case $\Delta_r \neq \Delta_b$ and case $\Delta_r = \Delta_b$.

- Case $\Delta_r \neq \Delta_b$. Suppose w.l.o.g. that $\Delta_r < \Delta_b$, and observe that since they have integer values it means $\Delta_r \leqslant \Delta_b - 1$. Now we show that, if we start from a configuration where the red opinion has the majority of nodes, the number of red nodes decreases in expectation. By using $\Delta_r \leqslant \Delta_b - 1$ in (67) we get

$$p(r) \leqslant \frac{c_r}{n^3} \left(c_r^2 + c_b(c_r - c_b)\Delta_b - c_r c_b + 3c_b^2\right). \tag{68}$$

If the majority of nodes is red then $c_r - c_b$ is positive, and since Δ_b can be at most 3 from (68) we get

$$p(r) \leqslant \frac{c_r}{n^3} \left(c_r^2 + 2c_r c_b\right). \tag{69}$$

Finally, if we put $c_r = n/2 + s$ and $c_b = n/2 - s$, for some positive s, in (69), we get that

$$p(r) \leqslant \frac{c_r}{n^3} \left(\frac{3}{4}n^2 + (n-s)s\right) \leqslant \frac{c_r}{n}. \tag{70}$$

- Case $\Delta_r = \Delta_b$. When $\Delta_r = \Delta_b$, observe that if the protocol is not a majority protocol then it must be $\Delta_r = \Delta_b \leqslant 2$. Hence, if we start again from a configuration where $c_r \geqslant c_b$, from (67) we get that

$$p(r) \leqslant \frac{c_r}{n^3} \left(c_r^2 + 2c_b(c_r - c_b) + 3c_b^2\right) = \frac{c_r}{n}. \tag{71}$$

In both cases, for any protocol $\mathcal{P}$ that does not behave like a majority protocol on triples with a clear majority, if we name X_t the random variable indicating the number of red nodes at round t, from (70) and (71) we get that $\mathbb{E}\left[X_{t+1} \mid X_t\right] \leqslant X_t$, hence X_t is a supermartingale. Now let τ be the random variable indicating the first time the chain hits one of the two absorbing states, i.e.

$$\tau = \inf\{t \in \mathbb{N} : X_t \in \{0, n\}\}.$$

Since $\Pr(\tau < \infty) = 1$ and all X_t's have values bounded between 0 and n, from the martingale stopping theorem[3] we get that $\mathbb{E}[X_\tau] \leqslant \mathbb{E}[X_0]$. If we

[3]See e.g. Chapter 17 in **[LPW09]** for a summary of martingales and related results.

start from a configuration that is $n/4$-imbalanced in favor of the red opinion, we have that $X_0 = n/2 + n/8$, and if we call ε is the probability that the process ends up with all blue nodes we have that $\mathbb{E}[X_\tau] = (1-\varepsilon)n$. Hence it must be $(1-\varepsilon)n \leqslant n/2 + n/8$ and the probability to end up with all blue nodes is $\varepsilon \geqslant 5/8 > 1/4$. Thus the protocol is not a $(n/4, 1/4)$-solver. ▣

LEMMA 24 (uniform property). *A constant $\eta > 0$ exists such that, if $\mathcal{P}$ is an $(\eta n, 1/4)$-solver, then its f must have the uniform property.*

PROOF. Thanks to the previous lemma, we can assume that f has the clear-majority property but a triple (r, g, b) exists such that $\delta_r < \max\{\delta_g, \delta_b\}$. Let us start the process with the following initial configuration having only the above 3 opinions and then show that the process does not converge to the plurality opinion r, w.h.p.:

$$\mathbf{c} = (c_r, c_g, c_b) = (n/3 + s,\ n/3,\ n/3 - s) \quad \text{where } s = \Theta(\sqrt{n \log n}).$$

We consider the "hardest" case where $\delta_r = 1$: the case $\delta_r = 0$ is simpler since in this case, no matter how the other $\delta' s$ are distributed, it is easy to see that the r.v. c_r decrease exponentially to 0 starting from the above configuration.

- Case $\delta_r = 1$, $\delta_g = 3$, and $\delta_b = 2$ (and symmetric cases). Starting from the above initial configuration, we can compute the probability

$$p(r) = \Pr\left(X_v = r \,\middle|\, C = \mathbf{c}\right)$$

that a node gets the opinion r.

$$\begin{aligned} p(r) &= \left(\frac{c_r}{n}\right)^3 + 3\left(\frac{c_r}{n}\right)^2 \frac{n - c_r}{n} + \frac{c_r c_g c_b}{n^3} \\ &= \frac{n+3s}{3n^3}\left(\left(\frac{n}{3}+s\right)^2 + 3\left(\frac{n}{3}+s\right)\left(\frac{2}{3}n - s\right) + \left(\frac{n}{3}\right)\left(\frac{n}{3} - s\right)\right). \end{aligned}$$

After some easy calculations, we get

$$p(r) = \frac{8}{27}\left(1 + O\left(\frac{s}{n}\right)\right).$$

As for $p(g)$, by similar calculations, we obtain the following bound

$$p(g) = \frac{10}{27}\left(1 - O\left(\frac{s^2}{n^2}\right)\right).$$

From the above two equations, we get the following bounds on the expectation of the r.v.'s X^r and X^g counting the nodes having opinion r and g, respectively (at the next round).

$$\begin{aligned} \mathbb{E}[X^r \mid \mathbf{C} = \mathbf{c}] &\leqslant \frac{8}{27}n + O(s) \ \text{ and} \\ \mathbb{E}[X^g \mid \mathbf{C} = \mathbf{c}] &\geqslant \frac{10}{27}n - O\left(\frac{s^2}{n}\right). \end{aligned}$$

By a standard application of the Chernoff bound (Lemma 76), we can prove that, if $s \leqslant \eta n$ for a sufficiently small $\eta > 0$, the initial value c_r decreases by a constant factor, w.h.p., going much below the new plurality c_g. Then, by applying iteratively the above reasoning we get that the process does not converge to r, w.h.p.

- Case $\delta_r = 1$, $\delta_g = 4$, and $\delta_b = 1$ (and symmetric cases). In this case it is even simpler to show that, starting from the same initial configuration considered in the previous case, the process does not converge to opinion r, w.h.p. ⊠

5.3.3. A lower bound for h-plurality

In Section 5.3.1, we have shown that the 3-Majority dynamics takes $\Theta(k \log n)$ rounds, w.h.p., to converge in the worst case. A natural question is whether by using the h-plurality protocol, with h slightly larger than 3, it is possible to significantly speed-up the process. We prove that this is not the case.

Let us consider a set of n nodes, each node having an opinion out of k possible ones. The h-plurality protocol works as follows:

> *At every round, every node picks h nodes uniformly at random (including herself and with repetitions) and updates her opinion according to the plurality of the opinions she sees (breaking ties u.a.r.)*

Let $j \in [k]$ be an arbitrary opinion, in the next lemma we prove that, if the number of nodes is smaller than $2n/k$ and if $k/h = \mathcal{O}(n^{(1-\varepsilon)/4})$, then the probability that the number of nodes having opinion j increases by a factor $(1 + h^2/k)$ is exponentially small.

LEMMA 25. *Let $\mathbf{c} = (c_1, \ldots, c_k)$ be a configuration and let $j \in [k]$ be an opinion such that $(n/k) \leqslant c_j \leqslant 2(n/k)$. If $k/h = \mathcal{O}(n^{(1-\varepsilon)/4})$ then it holds that*

$$\Pr\left(C_j^{(t+1)} \geqslant \left(1 + \frac{h^2}{k}\right) c_j \,\middle|\, \mathbf{C}^{(t)} = \mathbf{c} \right) \leqslant e^{-\Theta(n^{\varepsilon})}.$$

PROOF. Consider a specific node, say $u \in [n]$, let N_j be the number of nodes having opinion j picked by u during the sampling stage of the t-th round and let Y be the indicator random variable of the event that node u chooses opinion j at round $t+1$. We give an upper bound on the probability of the event $Y = 1$ by conditioning it on $N_j = 1$ and $N_j \geqslant 2$ (observe that if $N_j = 0$ node u cannot choose j as her opinion at the next round)

$$\Pr(Y_u = 1) \leqslant \Pr(Y_u = 1 \,|\, N_j = 1) \Pr(N_j = 1) + \Pr(N_j \geqslant 2). \tag{72}$$

Now observe that

- $\Pr(Y_u = 1 \,|\, N_j(u) = 1) \leqslant 1/h$ since it is exactly $1/h$ if all other sampled nodes have distinct opinions and it is 0 otherwise;
- $\Pr(N_j = 1) \leqslant hc_j/n$ since it can be bounded by the probability that at least one of the h samples gives opinion j;

- $\Pr(N_j \geqslant 2) \leqslant \binom{h}{2} c_j^2/n^2$ since it is the probability that a pair of sampled nodes exist with the same opinion j.

Hence, in (72) we have that

$$\Pr(Y=1) \leqslant \frac{c_j}{n} + \frac{h^2}{2} \cdot \frac{c_j^2}{n^2}.$$

Thus, for the expected number of nodes having opinion j at the next round we get

$$\mathbb{E}\left[C_j^{(t+1)} \mid \mathbf{C}^{(t)} = \mathbf{c}\right] \leqslant c_j + \frac{h^2}{2n} c_j^2 = c_j\left(1 + \frac{h^2}{2n} c_j\right) \leqslant c_j\left(1 + \frac{h^2}{k}\right),$$

where in the last inequality we used the hypothesis $c_j \leqslant 2(n/k)$. Since $C_j^{(t+1)}$ conditional on $\{\mathbf{C}^{(t)} = \mathbf{c}\}$ is a sum of n independent Bernoulli random variables, from the Chernoff bound (Lemma 76 with $\lambda_U = c_j h^2/k$), we finally get

$$\begin{aligned}\Pr\left(C_j^{(t+1)} \geqslant c_j\left(1 + 2\frac{h^2}{k}\right) \,\middle|\, \mathbf{C}^{(t)} = \mathbf{c}\right) &\leqslant \exp\left(-\frac{2(c_j h^2/k)^2}{n}\right) \\ &\leqslant \exp\left(-\Omega(n^{\varepsilon})\right),\end{aligned}$$

where in the last inequality we used $c_j \geqslant n/k$ and $k/h = \mathcal{O}(n^{(1-\varepsilon)/4})$. ∎

By adopting a similar argument to that used for proving Theorem 6, we can get a lower bound $\Omega(k/h^2)$ on the completion time of the h-plurality.

THEOREM 7 (Lower Bound for h-Majority). *Let $\mathbf{C}^{(t)}$ be the random variable indicating the configuration at round t according to the h-Plurality dynamics and let*

$$\tau = \inf\{t \in \mathbb{N} : \mathbf{C}^{(t)} \text{ is monochromatic}\}.$$

If the initial configuration $\mathbf{c} = (c_1, \ldots, c_k)$ is such that

$$\max\{c_j : j = 1, \ldots, k\} \leqslant \frac{3n}{2k},$$

then $\tau = \Omega(k/h^2)$ w.h.p.

PROOF. Since in the initial configuration for any opinion $j \in [k]$ we have that $c_j \leqslant 3n/(2k)$, from Lemma 25 it follows that the number of nodes supporting the plurality opinion increases at a rate smaller than $(1 + 2h^2/k)$ with probability exponentially close to 1. This easily implies a recursive relation of the form $C_j^{(t+1)} \leqslant \left(1 + 2h^2/k\right) C_j^{(t)}$ which, in turn, gives

$$C_j^{(t)} \leqslant \left(1 + \frac{2h^2}{k}\right)^t C_j^{(0)} \leqslant \left(1 + \frac{2h^2}{k}\right)^t \frac{3n}{2k}.$$

Thus, for $t < k/h^2 \log(4/3)$, we have that w.h.p.

$$C_j^{(t)} \leqslant \frac{3n}{2k}\left(1 + \frac{2h^2}{k}\right)^t < \frac{2n}{k},$$

concluding the proof. ☒

5.4. The 3-Majority Dynamics for Stabilizing Consensus

In this section, we move on to prove the results discussed in Section 2.2.2.

Since here we start investigating a different problem, in the following we partly recall some basic notation alredy introduced in Section 5.1. Since the communication graph is complete and nodes are anonymous, the overall system state at any round can be described by a *configuration* $\mathbf{c} := \big(c_1, ..., c_{|\Sigma|}\big)$, where the *support* c_i of opinion i is the number of nodes holding opinion i in that system's state. Given configuration $\mathbf{c}$, we say that an opinion i is *active* in $\mathbf{c}$ if $c_i > 0$ and, for any set of active opinions $W \subseteq \Sigma$, we define $\mathtt{m}(W) := \arg\min_{i \in W} c_i$. For any variable x of the process, we write $x^{(t)}$ if we are considering its value at round t and $X^{(t)}$ to denote the corresponding random variable.

The next lemma is an easy consequence of Lemma 17 and provides a general upper bound on the expected number of nodes supporting a given opinion at round $t+1$, given the configuration at round t.

LEMMA 26. *Let $\mathbf{c}$ be the configuration at round t and let $W \subseteq \Sigma$ be the subset of active opinions in $\mathbf{c}$. Then, for any opinion $i \in W$,*

$$\mathbb{E}\left[C_i^{(t+1)} \,\middle|\, \mathbf{C}^{(t)} = \mathbf{c}\right] \leqslant c_i \left(1 + \frac{c_i}{n} - \frac{1}{|W|}\right) \tag{73}$$

PROOF. From Lemma 17 we have

$$\begin{aligned}\mathbb{E}\left[C_i^{(t+1)} \,\middle|\, \mathbf{C}^{(t)} = \mathbf{c}\right] &= c_j \left(1 + \frac{c_j}{n} - \frac{1}{n^2}\sum_{h \in [k]} c_h^2\right) \\ &\leqslant c_i \left(1 + \frac{c_i}{n} - \frac{1}{|W|}\right),\end{aligned}$$

where in the inequality we used that the sum $\sum_{\ell \in W} c_\ell^2$ is minimized for $c_\ell = n/|W|$. ☒

Lemma 26 implies that opinions whose support falls below the average $n/|W|$ decrease in expectation. This expected drift is a key-ingredient of the analysis and, as we show in the next paragraph, it provides useful intuitions about the process. On the other hand, when $\mathbf{c}$ is almost uniform, the above *drift* turns out to be negligible and symmetry breaking is due to the inherent variance of the random process.

5.4.1. Approaches which don't seem to work

When the 3-Majority dynamics starts from configurations that exhibit a large initial support bias between the largest and the second-largest opinions, the approach adopted in Section 5.2 successfully exploits the fact that the initial plurality is preserved throughout the evolution of the random process, with an expected positive drift that is also preserved, w.h.p. An intuition of

this fact can be achieved from simple manipulations of (73). However, the aforementioned drift is only preserved if the largest opinion never changes, w.h.p., *no matter which the second-largest opinion is*: a condition that is not met by uniform configurations. A promising attempt to cope with uniform configurations is to consider the r.v.

$$S^{(t)} = C^{(t)}_{\mathbb{M}(t)} - C^{(t)}_{2\mathbb{M}(t)},$$

where $\mathbb{M}(t)$ and $2\mathbb{M}(t)$ are the r.v.s that take the index of (one of) the largest opinion and of (one of) the second-largest ones, respectively, in round t. For any *fixed* pair i, j, such that $c_i > c_j$, (73) implies that the difference $C_i^{(t+1)} - C_j^{(t+1)}$ in the next round is positive in expectation, so a suitable submartingale argument (similarly to those in [**LPW09**]), seemed to work in order to show that the system (rather quickly) achieves a "sufficiently-large" bias toward the plurality as to allow fast convergence. This approach would work if the *random* indices $\mathbb{M}$ and $2\mathbb{M}$ maintained their initial values across the entire duration of the process. Unfortunately, starting from uniform configurations, in the next round, the expected difference between the *new* largest opinion and the *new* second largest one may have no positive drift at all. Roughly speaking, in the next round, the r.v. $C^{(t+1)}_{2\mathbb{M}(t+1)}$ can be much larger than the r.v. $C^{(t+1)}_{\mathbb{M}(t)}$.

A promising dynamics for the stabilizing almost-consensus problem is the one introduced in [**DGM$^+$11**], in which nodes revise their opinions (assumed to be totally ordered) by taking the median between the currently held opinion and those held by two randomly sampled nodes. However, while we do not assume opinions to be integers (or totally ordered), their analysis strongly relies on the fact that the median opinion (or any good approximation of it) exhibits a strong increasing drift, even when starting from almost-uniform configuration, whereas no opinion is "special" to a majority rule when the starting configuration is uniform. The adoption of an inherently biased function as the median can have important consequences. To get an intuition, the reader may consider the following simple instance: $\Sigma = \{1, 2, 3\}$, with the system starting in configuration $c_1 = n/2, c_2 = 0, c_3 = n/2$ (Figure 11). At the end of the first round, a static adversary changes the values of $F = \log n$ nodes, equally distributed in c_1 and c_3, to value 2. The (non-valid) value 2 is the *global median* and standard counting arguments show that, while values 1 and 2 have no positive expected drift, the median has a strong expected drift that holds w.h.p. whenever $c_1, c_3 = \Theta(n)$. This might fool the system into the configuration in which $c_2 = n$, thus converging to a non-valid value.

5.4.2. The new approach

The analysis we present significantly departs from the above approaches. It is important to remark that, for $|\Sigma| \geqslant 3$, no previous analysis of the 3-Majority dynamics with almost-uniform initial configurations was known, even in the simpler non-adversarial case. On the other hand, while simpler,

the analysis of the non-adversarial case still has *per-se* interest and it requires to address some of the main technical challenges that also arise in the adversarial case. Section 5.4.3 is thus devoted to the analysis of the non-adversarial case, while an outline is given in the paragraphs that follow.

When the configuration is (approximately) uniform, Lemma 26 tells us that the process exhibits no significant drift toward any *fixed* opinion. Interestingly, things change if we consider the random variable $C_{\mathrm{m}}^{(t)}$, indicating the size of the smallest opinion supported at round t. Let $j \leqslant k$ be the number of active opinions in a given round t, we first prove that the expected value of $C_{\mathrm{m}}^{(t)}$ always exhibits a non-negligible negative drift:

$$\mathbb{E}\left[C_{\mathrm{m}}^{(t+1)} \mid \mathbf{C}^{(t)} = \hat{\mathbf{c}}\right] \leqslant \hat{c}_{\mathrm{m}} - \varepsilon \frac{\sqrt{n}}{j^{3/2}}\,, \tag{74}$$

for some constant $\varepsilon > 0$ (see Figure 17). This drift is essentially a consequence of Lemma 26 *and* of the standard deviation of r.v.s $C_i^{(t)}$s (see the proof of Lemma 28). The analysis then proceeds along consecutive phases, each consisting of a suitable number of consecutive rounds. If the number of active opinions at the beginning of the generic phase is j, we prove that, with positive constant probability, $C_{\mathrm{m}}^{(t)}$ vanishes within the end of the phase, so that the next phase begins with (at most) $j-1$ active opinions.

We clearly need a good bound on the length of a phase beginning with at most j opinions. To this aim, we derive a new upper bound - stated in Lemma 27 - on the *hitting time* of stochastic processes with expected drift that are defined by finite-state Markov chains [**LPW09**](74) to prove that, from any configuration with $j \leqslant k$ active opinions, $C_{\mathrm{m}}^{(t)}$ drops below the threshold $n/j - \sqrt{jn\log n}$ within $\mathcal{O}(\mathrm{poly}(j, \log n))$ rounds, with constant positive probability: This "hitting" event represents the exit condition from the symmetry-breaking stage of the phase. Indeed, once it occurs, we can consider *any fixed* active opinion i having support size c_i below the above threshold (thanks to the previous stage, we know that there is a good chance this opinion exists): We then show that C_i has a negative drift of order $\Omega(c_i/j)$. This allows us to prove that C_i drops from $n/j - \sqrt{jn\log n}$ to zero within $\mathcal{O}(\mathrm{poly}(j, \log n))$ further rounds, with positive constant probability. This interval of rounds is the dropping stage of the phase.

Ideally, the process proceeds along k consecutive phases, indexed as $j = k, k-1, \ldots, 2$, such that we are left with at most $j-1$ active opinions at the end of Phase j. In practice, we only have a constant probability that at least one opinion disappears during Phase j. However, using standard probabilistic arguments, we can prove that, w.h.p., for every j, the transition from j to $j-1$ active opinions takes a constant (amortized) number of phases, each requiring $\mathcal{O}(\mathrm{poly}(j, \log n))$ rounds.

The presence of a dynamic, adaptive adversary makes the above analysis technically more complex. A major issue is that a different definition of *phase* must be considered, since the adversary might permanently feed any

opinion so that the latter never dies. So the number of active opinions might not decrease from one phase to the next one. Essentially, we need to manage the persistence of "small" (valid or not) opinions: The end of a phase is now characterized by one "big" valid opinion that becomes "small" and, moreover, we need to show that, in general, "small" colors never become "big", no matter what the dynamic F-bounded adversary does. The dynamic-adversary case is described in Section 5.4.5.

5.4.3. Convergence Time without Adversary

Let $\mathcal{C} \subseteq \Sigma$ be the subset of valid opinions, i.e. those supported by at least one node in the initial configuration, and denote by $k = |\mathcal{C}|$ its size. This section is devoted to the proof of the following result, which is given in Section 5.4.4.

THEOREM 22 (Adversary-Free Upper Bound). *Starting from any initial configuration with $k \leqslant n^{1/3-\varepsilon}$ active opinions, where $\varepsilon > 0$ is an arbitrarily-small constant, the 3-Majority dynamics reaches consensus within $\mathcal{O}((k^2 \log^{1/2} n + k \log n) \cdot (k + \log n))$ rounds, w.h.p.*

We first provide the lemmas required for the process analysis and then we give the formal proof of the above theorem.

The next lemma shows an upper bound on the time it takes a stochastic process with values in $N = \{0, 1, \ldots, n\}$ to reach or exceed a target value m, under mild hypotheses on the process.

LEMMA 27. *Let $\{X_t\}_t$ be a Markov chain with finite state space Ω, let $f : \Omega \to N$ be a function mapping states of the chain to non-negative integer numbers, and let $\{Y_t\}_t$ be the stochastic process over N defined by $Y_t = f(X_t)$. Let $m \in N$ be a "target value" and let*

$$\tau = \inf\{t \in \mathbb{N} : Y_t \geqslant m\},$$

be the random variable indicating the first time Y_t reaches or exceeds value m. Assume that, for every state $x \in \Omega$ with $f(x) \leqslant m - 1$, it holds that

(1) (Positive drift). For some $\lambda > 0$

$$\mathbb{E}\left[Y_{t+1} \,|\, X_t = x\right] \geqslant f(x) + \lambda,$$

(2) (Bounded jumps). For some $\alpha > 1$

$$\Pr\left(Y_\tau \geqslant \alpha m \,|\, X_t = x\right) \leqslant \alpha m/n.$$

Then, for every starting state $x \in \Omega$, it holds that

$$\mathbb{E}\left[\tau \,|\, X_t = x\right] \leqslant 2\alpha \frac{m}{\lambda}.$$

IDEA OF PROOF. From Hypothesis *1* it follows that $Z_t = Y_t - \lambda t$ is a *submartingale* that satisfies the hypotheses of the Doob's *Optional Stopping*

Theorem (Theorem 27, pag. 272), thus

$$\begin{aligned} 0 \leqslant f(x) &= \mathbb{E}\left[Z_0 \mid X_t = x\right] \\ &\leqslant \mathbb{E}\left[Z_\tau \mid X_t = x\right] \\ &= \mathbb{E}\left[Y_\tau \mid X_t = x\right] - \lambda \mathbb{E}\left[\tau \mid X_t = x\right], \end{aligned}$$

and from Hypothesis *2* it follows that $\mathbb{E}\left[Y_\tau \mid X_t = x\right] \leqslant 2\alpha m$. ▣

Proof. Consider the stochastic process $Z_t = Y_t - \lambda t$ and observe that for any state $x \in \Omega$ with $f(x) \leqslant m - 1$ it holds that

$$\begin{aligned} \mathbb{E}\left[Z_{t+1} \mid X_t = x\right] &= \mathbb{E}\left[Y_{t+1} \mid X_t = x\right] - \lambda(t+1) \\ &\geqslant f(x) + \lambda - \lambda(t+1) \\ &\geqslant f(x) - \lambda t, \end{aligned}$$

where in the inequality we used Hypotheses *1*. Thus Z_t is a *submartingale* up to the stopping time τ, i.e. $\mathbb{E}\left[Z_{t+1} \mid X_t\right] \geqslant Z_t$ for any $t < \tau$. Moreover, since $|Y_t| \leqslant n$ the *jumps* of Z_t can be bounded by a value independent of t

$$|Z_{t+1} - Z_t| = |Y_{t+1} - \lambda(t+1) - Y_t + \lambda t| \leqslant n + \lambda.$$

It is also easy to see that Hypothesis *1* implies $\mathbb{E}\left[\tau \mid X_t = x\right] < \infty$. Thus, we can apply *Doob's Optional Stopping Theorem* (Theorem 27, pag. 272). It then follows that

$$\mathbb{E}\left[Z_\tau \mid X_t = x\right] \geqslant \mathbb{E}\left[Z_0 \mid X_t = x\right] = f(x)$$

and, since

$$\mathbb{E}\left[Z_\tau \mid X_t = x\right] = \mathbb{E}\left[Y_\tau \mid X_t = x\right] - \lambda \mathbb{E}\left[\tau \mid X_t = x\right],$$

we have that

$$\mathbb{E}\left[\tau \mid X_t = x\right] \leqslant \frac{\mathbb{E}\left[Y_\tau \mid X_t = x\right] - f(x)}{\lambda} \leqslant \frac{\mathbb{E}\left[Y_\tau \mid X_t = x\right]}{\lambda}.$$

Finally, we get

$$\begin{aligned} &\mathbb{E}\left[Y_\tau \mid X_t = 0\right] \\ &= \sum_{j=1}^{n} j \Pr\left(Y_\tau = j \mid X_t = 0\right) \\ &= \sum_{j=1}^{\lfloor \alpha m \rfloor} j \Pr\left(Y_\tau = j \mid X_t = 0\right) + \sum_{j=\lfloor \alpha m \rfloor + 1}^{n} j \Pr\left(Y_\tau = j \mid X_t = 0\right) \\ &\leqslant (\alpha m) + n \Pr\left(Y_\tau > \alpha m \mid X_t = 0\right) \leqslant 2\,(\alpha m), \end{aligned}$$

where in the last inequality we used Hypothesis *2*. ▣

We next use the above lemma to bound the time required by the *symmetry-breaking* stage.

LEMMA 28 (Symmetry-Breaking Stage). *Let* $\mathbf{c}$ *be any configuration with* j *active opinions. Within* $t = \mathcal{O}\left(j^2 \log^{1/2} n\right)$ *rounds it holds that*

$$\Pr\left(\exists i \,:\, C_i^{(t)} \leqslant n/j - \sqrt{jn \log n} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right) \geqslant \frac{1}{2}.$$

PROOF. Let J be the set of j active opinions in $\mathbf{c}$ and let

$$\mathbf{C}^{(t)} = \left(C_i^{(t)} \,:\, i \in J\right),$$

be the random variable indicating the opinion configuration at round t, where we assume $\mathbf{C}^{(0)} = \mathbf{c}$. Let

$$C_{\mathrm{m}}^{(t)} = \min\left\{C_i^{(t)} \,:\, i \in J\right\},$$

be the minimum among all $C_i^{(t)}$s and consider the stochastic process $\{Y_t\}_t$ defined as $Y_t = \lfloor n/j \rfloor - C_{\mathrm{m}}^{(t)}$. Observe that Y_t takes values in $\{0, 1, \ldots, \lfloor n/j \rfloor\}$ and it is a function of $\mathbf{C}^{(t)}$. We are interested in the first time Y_t becomes at least as large as $\sqrt{jn \log n}$, i.e.

$$\tau = \inf\left\{t \in \mathbb{N} \,:\, Y_t \geqslant \sqrt{jn \log n}\right\}.$$

We now show that $\{Y_t\}_t$ satisfies Hypothesis *1* and *2* of Lemma 27, with $\lambda = \varepsilon\sqrt{n}/j^{3/2}$, for a suitable constant $\varepsilon > 0$.

1. Let $\hat{\mathbf{c}} = (\hat{c}_i \,:\, i \in J)$ be any configuration with j active opinions such that $\hat{c}_{\mathrm{m}} > n/j - \sqrt{jn \log n}$. We want to prove that

$$\mathbb{E}\left[C_{\mathrm{m}}^{(t+1)} \,|\, \mathbf{C}^{(t)} = \hat{\mathbf{c}}\right] \leqslant c_{\mathrm{m}} - \varepsilon \frac{\sqrt{n}}{j^{3/2}}. \tag{75}$$

Two cases may arise.

Case $\hat{c}_{\mathrm{m}} > n/j - 2\varepsilon\sqrt{n/j}$**.** Observe that, in this case, the r.v.s $\{C_i^{(t+1)} \,:\, i \in J\}$ conditional on $\{\mathbf{C}^{(t)} = \hat{\mathbf{c}}\}$ have standard deviation $\Omega(\sqrt{n/j})$. Moreover, they are binomial and negatively associated. Hence, by choosing ε small enough, from the Central Limit Theorem we have that

$$\Pr\left(\exists i \in J \,:\, C_i^{(t+1)} \leqslant \frac{n}{j} - 6\varepsilon \cdot \sqrt{\frac{n}{j}} \,|\, \mathbf{C}^{(t)} = \hat{\mathbf{c}}\right) \geqslant \frac{1}{2}.$$

We thus get

$$\begin{aligned}\mathbb{E}\left[C_{\mathrm{m}}^{(t+1)} \,|\, \mathbf{C}^{(t)} = \hat{\mathbf{c}}\right] &\leqslant \frac{1}{2}\left(\frac{n}{j} - 6\varepsilon \cdot \sqrt{\frac{n}{j}}\right) + \frac{1}{2} \cdot \frac{n}{j} \\ &= \frac{n}{j} - 3\varepsilon\sqrt{\frac{n}{j}} \leqslant c_{\mathrm{m}} - \varepsilon\sqrt{\frac{n}{j}} \leqslant c_{\mathrm{m}} - \varepsilon\frac{\sqrt{n}}{j^{3/2}}.\end{aligned}$$

Case $\hat{c}_{\mathrm{m}} \leqslant n/j - 2\varepsilon\sqrt{n/j}$. (75) easily follows from Lemma 26. Indeed, let $i \in J$ be an opinion such that $\hat{c}_i = \hat{c}_{\mathrm{m}}$, then

$$\begin{aligned}(76)\qquad \mathbb{E}\left[C_{\mathrm{m}}^{(t+1)} \,|\, \mathbf{C}^{(t)} = \hat{\mathbf{c}}\right] &\leqslant \mathbb{E}\left[C_i^{(t+1)} \,|\, \mathbf{C}^{(t)} = \hat{\mathbf{c}}\right] \\ &\leqslant \hat{c}_i\left(1 + \frac{\hat{c}_i}{n} - \frac{1}{j}\right) \\ &\leqslant \hat{c}_i\left(1 - \frac{2\varepsilon}{\sqrt{nj}}\right) \\ &\leqslant \hat{c}_i - \frac{\varepsilon\sqrt{n}}{j^{3/2}} = \hat{c}_{\mathrm{m}} - \varepsilon\frac{\sqrt{n}}{j^{3/2}}\end{aligned}$$

where we used the case's condition and the fact that $\hat{c}_i = \hat{c}_{\mathrm{m}} \geqslant n/(2j)$.

2. Since the random variables $\{C_i^{(t+1)} : i \in J\}$ are binomial, conditional on the configuration at round t, it is possible to apply the Chernoff bound (Lemma 76, though with some care) to prove that

$$(77)\qquad \Pr\left(Y_\tau \geqslant \alpha\sqrt{jn\log n} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right) \leqslant \frac{1}{n},$$

for some constant $\alpha > 1$. Though this result seems intuitive, its formal proof is less obvious, since τ is a stopping time and thus itself a random variable. Lemma 35 in Section 5.4.6 offers a formal proof of the above statement.

From (75) and (77), we have that $\{Y_t\}_t$ satisfies the hypotheses of Lemma 27 with $m = \sqrt{jn\log n}$ and $\lambda = \varepsilon\sqrt{n}/j^{3/2}$. Hence

$$\mathbb{E}\left[\tau \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right] < j^2\sqrt{\log n}$$

and, from Markov inequality, for $t = 2j^2\sqrt{\log n}$, we finally get

$$\begin{aligned}&\Pr\left(\forall\, i \in J : C_i^{(t)} \geqslant n/j - \sqrt{jn\log n} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right) \\ &\leqslant \Pr\left(\tau > 2j^2\sqrt{\log n} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right) \leqslant \frac{1}{2}.\end{aligned}$$

gg

5.4.4. The Survival of the Bigger

We now provide the analysis of the *dropping* stage: More precisely, we show that, if the system starts with up to j active opinions and one of them (say i) is below the threshold $n/j - \sqrt{jn\log n}$, then i drops to the smaller threshold $j^2\log n$ within $\mathcal{O}(j\log n)$ additional rounds. This bound can be proved w.h.p. since, in this regime, C_i is still sufficiently large to apply the Chernoff bound. This concentration result is not necessary to the purpose of proving Theorem 22, while it is a key ingredient in the analysis of the adversarial case (Theorem 8). The next lemma can be proved by standard concentration arguments - applied in an iterative way - on the r.v. $C_i^{(t)}$ (see Section 5.4.6).

LEMMA 29 (Dropping Stage 1). *Let* $\mathbf{c}$ *be any configuration with* $j \leqslant n^{1/3-\varepsilon}$ *active opinions, where* $\varepsilon > 0$ *is an arbitrarily-small positive constant, and such that an opinion* i *exists with* $c_i \leqslant n/j - \sqrt{jn\log n}$. *Within* $t = \mathcal{O}(j\log n)$ *rounds opinion* i *becomes* $\mathcal{O}\left(j^2\log n\right)$, *w.h.p.*

In the next lemma we prove that once c_i becomes smaller than $n/(2j)$, then opinion i disappears within further $\mathcal{O}(j\log n)$ rounds with constant probability. We only give an outline of the proof (the full proof is presented in Section 5.4.6).

LEMMA 30 (Dropping Stage 2). *Let* $\mathbf{c}$ *be any configuration with* $j \leqslant n^{1/3-\varepsilon}$ *active opinions, where* $\varepsilon > 0$ *is an arbitrarily-small positive constant, and such that an opinion* i *exists with* $c_i \leqslant n/(2j)$. *Within* $t = \mathcal{O}(j\log n)$ *rounds opinion* i *disappears with probability at least* $1/2$.

IDEA OF PROOF. If $c_i \leqslant n/(2j)$ in configuration $\mathbf{c}$, then from Lemma 26 it follows that

$$\mathbb{E}\left[C_i^{(t+1)} \,|\, \mathbf{C}^{(t)} = \mathbf{c}\right] \leqslant c_i\left(1 - \frac{1}{2j}\right)$$

Moreover, since $C_i^{(t+1)}$ conditional on $\left\{\mathbf{C}^{(t)} = \mathbf{c}\right\}$ is binomial, if $j \leqslant n^{1/3-\varepsilon}$, from the Chernoff bound (Lemma 76) it follows that

$$\Pr\left(C_i^{(t+1)} > n/(2j) \,|\, \mathbf{C}^{(t)} = \mathbf{c}\right) \leqslant e^{-\Theta(n^{\varepsilon})}.$$

Hence, it is easy to check that for any initial configuration $\mathbf{c}$ with $c_i \leqslant n/(2j)$ the following recursive relation holds

$$\mathbb{E}\left[C_i^{(t)} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right] \leqslant \left(1 - \frac{1}{2j}\right)\mathbb{E}\left[C_i^{(t-1)} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right] + e^{-n^{\varepsilon/2}}$$

that for some $t = \mathcal{O}(j\log n)$ gives $\mathbb{E}\left[C_i^{(t)} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right] \leqslant 1/2$. Since $C_i^{(t)}$ is a non-negative integer-valued r.v., the thesis then follows from the Markov inequality. ▫

Armed with lemmas 29 and 30, we are ready to prove Theorem 22.

THEOREM 22 (Adversary-Free Upper Bound). *Starting from any initial configuration with* $k \leqslant n^{1/3-\varepsilon}$ *active opinions, where* $\varepsilon > 0$ *is an arbitrarily-small constant, the 3-Majority dynamics reaches consensus within* $\mathcal{O}((k^2\log^{1/2} n + k\log n)\cdot(k+\log n))$ *rounds, w.h.p.*

PROOF OF THEOREM 22. From Lemmas 28, 29, and 30 it follows that from any configuration with $j \leqslant k$ active opinions, within $\mathcal{O}(k^2\sqrt{\log n} + k\log n)$ rounds at least one of the opinions disappears with probability at least $1/4$. Thus, within $\mathcal{O}((k^2\sqrt{\log n}+k\log n)(k+\log n))$ rounds, all opinions but one disappear, w.h.p. ▫

5.4.5. Convergence Time with Adversary

In this section we consider the presence of a Byzantine adversary that can adaptively change the opinions of a subset of nodes in order to

(i) delay the convergence time toward a valid consensus, or
(ii) make the system converge to a non-valid opinion.

We consider two different kinds of adversaries: A static one and a stronger, dynamic one.

DEFINITION 15 (F-static adversary). Let $\mathbf{c}$ be the initial configuration: At the beginning of the process the adversary looks at $\mathbf{c}$ and can replace the opinions of at most $F = n/k - \sqrt{kn\log n}$ nodes with arbitrary opinions in Σ. Then, the adversary is not allowed to perform any further action during the execution of the protocol.

We consider the case $F = n/k - \sqrt{kn\log n}$. Since any opinion the adversary may introduce has size less than $n/k - \sqrt{kn\log n}$, as a simple consequence of the dropping stage (see Lemmas 29 and 30), the static adversarial case easily reduces to the non-adversarial one. We thus get the following result.

COROLLARY 6 (Upper Bound with Static-Adversary). *Starting from any initial configuration with $k \leqslant n^{\alpha}$ active opinions, where $\alpha > 0$ is a suitable constant, the 3-Majority dynamics reaches almost-consensus within $\mathcal{O}((k^2\sqrt{\log n} + k\log n)\cdot(k+\log n))$ rounds, in the presence of any F-static adversary with $F \leqslant n/k - \sqrt{kn\log n}$, w.h.p.*

We now define the actions of an F-dynamic adversary over the studied process can be described as follows.

DEFINITION 16 (F-Dynamic Adversary). At every round t, after nodes have updated their opinions (i.e. once the configuration $\mathbf{C}^{(t)} = \mathbf{c}^{(t)}$ is realized), the F-dynamic adversary looks at the current configuration and replaces the opinion of up to F nodes with any opinion in Σ. We define $\tilde{\mathbf{C}}^{(t)}$ as the configuration that results from the adversary's action on $\mathbf{c}^{(t)}$ and

$$D_i^{(t)} = D_i^{(t)}(\mathbf{c}^{(0)}, \tilde{\mathbf{c}}^{(0)}, \ldots, \mathbf{c}^{(t-1)}, \tilde{\mathbf{c}}^{(t-1)}, \mathbf{c}^{(t)}),$$

as the r.v. corresponding to the number of nodes that the adversary adds or removes from c_i (note that $\sum_{i\in\Sigma}|D_i| \leqslant 2F$) at the end of the t-th round, based on all the past history of the process, i.e.

$$\tilde{\mathbf{C}}^{(t)} = \left(C_1^{(t)} + D_1^{(t)}, \ldots, C_{|\Sigma|}^{(t)} + D_{|\Sigma|}^{(t)}\right).$$

In what follows we consider an F-dynamic adversary with $F \leqslant \beta\sqrt{n}/(k^{\frac{5}{2}}\log n)$ for a suitable positive constant β. As we show in the proof of Lemma 31, this bound on F turns out to be almost tight if the goal is to converge to an almost-consensus regime in polynomial time, w.h.p.

The presence of the adversary requires us to distinguish between valid and non valid opinions. So, we recall that the set of valid opinions $\mathcal{C} \subseteq \Sigma$ is

the subset of active opinions in the initial configuration and, in the sequel, we denote k as the number of valid opinions, i.e., $k := |\mathcal{C}|$ and define $\bar{\mathcal{C}} := \Sigma - \mathcal{C}$.

We are now ready to state our main result in the presence of the dynamic adversary.

THEOREM 8 (Upper Bound with Dynamic-Adversary). *Let $k \leqslant n^{\alpha}$ and $F \leqslant \beta\sqrt{n}/(k^{\frac{5}{2}} \log n)$ for some constants $\beta, \alpha > 0$. The 3-Majority dynamics is a stabilizing almost-consensus protocol in the presence of any F-dynamic adversary and its convergence time is $\mathcal{O}((k^2\sqrt{\log n} + k \log n)(k + \log n))$, w.h.p.*

In order to prove the above theorem, we need to "improve" the technical lemmas shown in the previous section for the non-adversarial case. Informally speaking, the adversary can introduce "small" non-valid opinions and it can keep small valid opinions active that, we know, they would otherwise disappear. These facts lead us to the problem of managing "small" opinions.

PROOF OF THEOREM 8. The rigorous definition of "small opinion" is determined by the minimal negative drift for $C_{\mathrm{m}}^{(t)}$ we derived in the proof of Lemma 28 (see Section (76)).

DEFINITION 17 (Small Opinions). Let $S := \{i \mid c_i \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}\}$ be the set of the *small opinions*, where γ is some constant such that $\gamma > \beta$, and let its complement $B := \bar{S} = \{i \mid c_i > \gamma\sqrt{n}/k^{\frac{3}{2}}\}$ be the set of the *big opinions*.

It turns out that we cannot define the end of a phase as we did in the non-adversarial case, namely, at least one (valid) opinion dies. We rather assume that, without loss of generality, all k valid opinions are big when the process begins. The consequent new definition of a phase is the following: phase j is an interval of consecutive rounds, in each of which exactly j big valid opinions are present. The goal then is to show that at the end of phase j, one of the j initially big opinions becomes small and, moreover, this opinion (and no other small opinion) never gets big again.

In the symmetry-breaking stage of each phase, we thus need to show that the negative drift of $C_{\mathrm{m}}^{(t)}$ (notice that the latter now denotes the minimum among the j big opinions) cannot be opposed by the actions of the F-dynamic adversary, provided that $F \leqslant \beta\sqrt{n}/(k^{\frac{5}{2}} \log n)$.

LEMMA 31 (Symmetry-Breaking Stage with Adversary). *Let $\tilde{\mathbf{c}}$ be any configuration such that $|B| = j$ and $\sum_{i\in\bar{\mathcal{C}}} \tilde{c}_i \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$. Within $t = \mathcal{O}(j^2 \log^{1/2} n)$ rounds, with probability at least $1/2$ it holds that*

i) $|B| = j$, $\sum_{i\in\bar{\mathcal{C}}} \tilde{C}_i \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$, and
ii) there exists an $i \in B^{(t)}$ such that $\tilde{C}_i^{(t)} \leqslant n/j - \sqrt{jn \log n}$.

The formal proof of the above lemma is given in Section 5.4.6. Informally, the proof is obtained via two different technical steps:

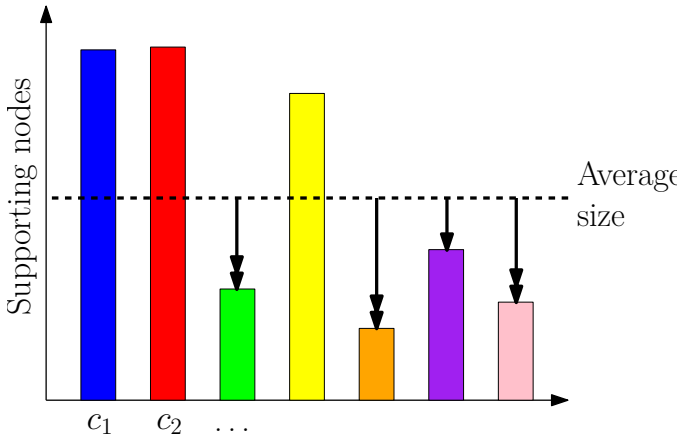

FIGURE 17. The negative drift affects any color whose size is smaller than the average size of the colors which are still present in the system.

i) The bound on the expected negative drift for $C_{\mathrm{m}}^{(t)}$ given by the following Lemma 32, which considers both the presence of small good opinions and the adversary's opposing action (for its proof see Section 5.4.6).
ii) A novel use of Lemma 27 on the hitting time of random processes in order to bound the expect time of the symmetry-breaking stage. We in fact need to define a new stopping condition that also includes some "bad" event: Some small (valid or not) opinion become big. More precisely, in Lemma 33 (its formal proof can be found in Section 5.4.6), we prove the following key-properties of the process in the presence of the dynamic adversary:
   (1) if in a given round a valid opinion is small then it keeps small in the following round, w.h.p., i.e. $S^{(t-1)} \subseteq S^{(t)}$;
   (2) the size of the overall set of non valid opinions stays below $\gamma\sqrt{n}/k^{\frac{3}{2}}$, w.h.p., i.e. $\sum_{i\in\bar{\mathcal{C}}} c_i \geqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$.

LEMMA 32 (Dropping Stage 1 with Adversary). *Let $\tilde{\mathbf{c}}$ be any configuration such that $|B| \leqslant j$ and $\sum_{i\in\bar{\mathcal{C}}} \tilde{c}_i^{(t)} \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$. For some constant $\alpha > 0$, for any opinion i such that $\tilde{c}_i \geqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$, it holds*

$$\mathbb{E}\left[C_i^{(t+1)} \,\middle|\, \tilde{\mathbf{C}}^{(t)} = \tilde{\mathbf{c}}\right] \leqslant \tilde{c}_i\left(1 - \frac{1}{j} + \frac{\tilde{c}_i + \alpha\sqrt{n/k}}{n}\right)$$

$$\mathbb{E}\left[\tilde{C}_i^{(t+1)} \,\middle|\, \tilde{\mathbf{C}}^{(t)} = \tilde{\mathbf{c}}\right] \leqslant \tilde{c}_i(1 - \eta(i,j)), \text{ where} \tag{78}$$

$$\eta(i,j) = \min\left\{\frac{1}{j} - \frac{\tilde{c}_i + \alpha\sqrt{n/k}}{n}\ ,\ \frac{1}{2}\left(\frac{1}{j} - \frac{\tilde{c}_i}{n}\right)\right\}$$

LEMMA 33 (Small Stays Small). *If* $\tilde{\mathbf{c}}^{(t)}$ *is such that* $\sum_{i\in\bar{\mathcal{C}}} \tilde{c}_i^{(t)} \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$, *then* $\sum_{i\in\bar{\mathcal{C}}} \tilde{C}_i^{(t+1)} \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$ *and* $S^{(t)} \subseteq S^{(t+1)}$, *w.h.p.*

The dropping stage of phase j is now defined as the interval of rounds in which $C_{\mathrm{m}}^{(t)}$ drops from the symmetry-breaking threshold $n/j - \sqrt{jn\log n}$ to the size of small opinions i.e. $\gamma\sqrt{n}/k^{\frac{3}{2}}$. Similarly to the non-adversarial case, we can here fix the big opinion i that is dropped below the symmetry-breaking threshold and look at its negative drift derived in Lemma 32. The drift turns out to be strong enough to compensate the possible actions of any F-bounded adversary and implies an $O(j\log n)$ bound on the time required by this second stage of phase j. This result is stated in the following Lemma (its proof is given in Section 5.4.6).

LEMMA 34 (Dropping Stage 2 with Adversary). *Assume that, at round* t', $\tilde{\mathbf{c}}^{(t')}$ *is such that*

- $\sum_{i\in\bar{\mathcal{C}}} c_i \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$,
- $|B^{(t')}| = j$, *and*
- *an* $i \in B^{(t')}$ *exists such that*

$$\gamma\sqrt{n}/k^{\frac{3}{2}} \leqslant c_i^{(t')} \leqslant n/j - \sqrt{kn\log n}.$$

Then, a round $t'' = t' + O(k\log n)$ *exists such that w.h.p.*

- $\sum_{i\in\bar{\mathcal{C}}} \tilde{C}_i^{(t'')} \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$,
- $i \in S^{(t'')}$ *and*
- $|B^{(t'')}| \leqslant j-1$.

Finally, after k phases, we are left with one (valid) opinion that accounts for $n - O(\sqrt{n})$ nodes, while the remaining nodes can have any (possibly non valid) opinion and reflect the presence of the adversary. In fact, this is what happens with high probability. gg

5.4.6. Technical lemmas of the analysis

In this section we complete the proof of Theorem 8 by proving (77) and the lemmas 31, 32, 34 and 33. Throughout the section, recall that we assume $F \leqslant \beta\sqrt{n}/(k^{\frac{5}{2}}\log n)$. We first provide a formal proof for (77).

LEMMA 35. *Let* $\mathbf{c}$ *be any configuration with* j *active opinions. Consider the stochastic process* $\{Y_t\}_t$ *defined as* $Y_t = \left\lfloor \frac{n}{j} \right\rfloor - C_{\mathrm{m}}^{(t)}$ *and define the stopping time* $\tau = \inf\left\{t\in\mathbb{N} : Y_t \geqslant \sqrt{jn\log n}\right\}$. *Then*

$$\Pr\left(Y_\tau > \alpha\sqrt{jn\log n} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right) \leqslant \frac{1}{n},$$

for some constant $\alpha > 1$.

PROOF. Observe that Y_τ is well defined, because $\mathbb{E}\left[\tau \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right] < \infty$ as a consequence of the fact that $C_{\mathrm{m}}^{(t)}$ has a negative drift (see the proof of Lemma 28).

From the definition of Y_t, we have

$$\begin{aligned}
&\Pr\left(Y_\tau > \alpha\sqrt{jn\log n}\,|\,\mathbf{C}^{(0)}=\mathbf{c}\right)\\
&=\Pr\left(C_{\mathtt{m}}^{(\tau)} < \left\lfloor\frac{n}{j}\right\rfloor - \alpha\sqrt{jn\log n}\,|\,\mathbf{C}^{(0)}=\mathbf{c}\right)\\
&=\Pr\left(\exists\ell : C_{\ell}^{(\tau)} < \left\lfloor\frac{n}{j}\right\rfloor - \alpha\sqrt{jn\log n}\,|\,\mathbf{C}^{(0)}=\mathbf{c}\right)\\
&\leqslant \sum_{\ell=1}^{j}\Pr\left(C_{\ell}^{(\tau)} < \left\lfloor\frac{n}{j}\right\rfloor - \alpha\sqrt{jn\log n}\,|\,\mathbf{C}^{(0)}=\mathbf{c}\right). \qquad (79)
\end{aligned}$$

To prove the lemma, we prove that each term in (79) is upper bounded by n^{-2}, by chossing α large enough.

Given any opinion ℓ, any comparison operator $\odot \in \{<, \leqslant, \geqslant, >\}$ and any round t, let

$$\mathcal{E}_{\ell\odot}^{(t)} = \text{“}C_{\ell}^{(t)} \odot \left\lfloor\frac{n}{j}\right\rfloor - \sqrt{jn\log n}\text{”}.$$

From the definition of the stopping time τ, for any opinion ℓ we have

$$\begin{aligned}
&\Pr\left(C_{\ell}^{(\tau)} < \left\lfloor\frac{n}{j}\right\rfloor - \alpha\sqrt{jn\log n}\,|\,\mathbf{C}^{(0)}=\mathbf{c}\right)\\
&=\sum_{t=1}^{\infty}\Pr\left(\left(C_{\ell}^{(t)} < \left\lfloor\frac{n}{j}\right\rfloor - \alpha\sqrt{jn\log n}\right)\bigwedge(\tau=t)\,|\,\mathbf{C}^{(0)}=\mathbf{c}\right)\\
&=\sum_{t=1}^{\infty}\Pr\left(\left(C_{\ell}^{(t)} < \left\lfloor\frac{n}{j}\right\rfloor - \alpha\sqrt{jn\log n}\right)\bigwedge\mathcal{E}_{\mathtt{m}\leqslant}^{(t)}\,\Big|\,\bigwedge_{s=1}^{t-1}\mathcal{E}_{\mathtt{m}>}^{(s)}\wedge(\mathbf{C}^{(0)}=\mathbf{c})\right)\\
&\qquad\cdot\Pr\left(\bigwedge_{s=1}^{t-1}\mathcal{E}_{\mathtt{m}>}^{(s)}\,|\,\mathbf{C}^{(0)}=\mathbf{c}\right)\\
&=\sum_{t=1}^{\infty}\Pr\left(C_{\ell}^{(t)} < \left\lfloor\frac{n}{j}\right\rfloor - \alpha\sqrt{jn\log n}\,\Big|\,\bigwedge_{s=1}^{t-1}\mathcal{E}_{\mathtt{m}>}^{(s)}\wedge(\mathbf{C}^{(0)}=\mathbf{c})\right)\\
&\qquad\cdot\Pr\left(\bigwedge_{s=1}^{t-1}\mathcal{E}_{\mathtt{m}>}^{(s)}\,|\,\mathbf{C}^{(0)}=\mathbf{c}\right), \qquad (80)
\end{aligned}$$

where the last equality follows from the fact that

$$C_{\ell}^{(t)} < \left\lfloor\frac{n}{j}\right\rfloor - \alpha\sqrt{jn\log n},$$

implies

$$C_{\mathtt{m}}^{(t)} < \left\lfloor\frac{n}{j}\right\rfloor - \sqrt{jn\log n}.$$

We next focus on bounding the term

$$\Pr\left(\bigwedge_{s=1}^{t-1} \mathcal{E}^{(s)}_{\mathtt{m}>} \mid \mathbf{C}^{(0)} = \mathbf{c}\right) \tag{81}$$

in (80). We can write

$$\begin{aligned}\Pr\left(\bigwedge_{s=1}^{t-1} \mathcal{E}^{(s)}_{\mathtt{m}>} \mid \mathbf{C}^{(0)} = \mathbf{c}\right) &= \prod_{s=1}^{t-1} \Pr\Big(\mathcal{E}^{(s)}_{\mathtt{m}>} \,\Big|\, \bigwedge_{r=1}^{s-1} \mathcal{E}^{(r)}_{\mathtt{m}>} \wedge \mathbf{C}^{(0)} = \mathbf{c}\Big) \\ &= \prod_{s=1}^{t-1} \Pr\left(\mathcal{E}^{(s)}_{\mathtt{m}>} \,\middle|\, \mathcal{E}^{(s-1)}_{\mathtt{m}>} \wedge \mathbf{C}^{(0)} = \mathbf{c}\right),\end{aligned}$$

where the last equality follows since the process of the 3-Majority dynamics is Markovian. We can upper bound

$$\begin{aligned}&\Pr\left(\mathcal{E}^{(s)}_{\mathtt{m}>} \,\middle|\, \mathcal{E}^{(s-1)}_{\mathtt{m}>} \wedge \mathbf{C}^{(0)} = \mathbf{c}\right) \\ &= \sum_{\hat{\mathbf{c}} \in \mathcal{S}_{\mathtt{m}}} \Pr\left(\mathcal{E}^{(s)}_{\mathtt{m}>} \mid \mathbf{C}^{(s-1)} = \hat{\mathbf{c}}\right) \cdot \Pr\left(\mathbf{C}^{(s-1)} = \hat{\mathbf{c}} \,\middle|\, \mathcal{E}^{(s-1)}_{\mathtt{m}>} \wedge \mathbf{C}^{(0)} = \mathbf{c}\right) \\ &\leqslant \sum_{\hat{\mathbf{c}} \in \mathcal{S}_{\mathtt{m}}} \Pr\left(\mathcal{E}^{(s)}_{\bar{m}>} \mid \mathbf{C}^{(s-1)} = \hat{\mathbf{c}}\right) \cdot \Pr\left(\mathbf{C}^{(s-1)} = \hat{\mathbf{c}} \,\middle|\, \mathcal{E}^{(s-1)}_{\mathtt{m}>} \wedge \mathbf{C}^{(0)} = \mathbf{c}\right),\end{aligned}$$

where $\bar{m}$ is the value of $\mathtt{m}$ at time $s-1$ (breaking ties arbitrarily), and $\mathcal{S}_{\mathtt{m}}$ is the set of possible configurations which realize $\mathcal{E}^{(s-1)}_{\mathtt{m}>}$, that is

$$\mathcal{S}_{\mathtt{m}} := \left\{\hat{\mathbf{c}} \,:\, \hat{c}_{\mathtt{m}} > \left\lfloor \frac{n}{j} \right\rfloor - \sqrt{jn \log n}\right\}.$$

We can also upper bound $\Pr(\mathcal{E}^{(s)}_{\bar{m}>} \mid \mathbf{C}^{(s-1)} = \hat{\mathbf{c}})$ by using a "reverse" Chernoff bound[4] (Theorem 26). In particular, for a suitable constant β it is possible to show that

$$\begin{aligned}&\Pr\left(C^{(s)}_{\bar{m}} > (1-\delta)\mathbb{E}\left[C^{(s)}_{\bar{m}} \mid \mathbf{C}^{(s-1)} = \hat{\mathbf{c}}\right] \mid \mathbf{C}^{(s-1)} = \hat{\mathbf{c}}\right) \\ &\leqslant 1 - e^{-\beta\delta^2 \mathbb{E}\left[C^{(s)}_{\bar{m}} \mid \mathbf{C}^{(s-1)} = \hat{\mathbf{c}}\right]}.\end{aligned}$$

By choosing

$$\delta = \frac{\sqrt{jn \log n}}{\mathbb{E}\left[C^{(s)}_{\bar{m}} \mid \mathbf{C}^{(s-1)} = \hat{\mathbf{c}}\right]}$$

and noting that

$$\frac{n}{2j} \leqslant \mathbb{E}\left[C^{(s)}_{\bar{m}} \mid \mathbf{C}^{(s-1)} = \hat{\mathbf{c}}\right] \leqslant \frac{n}{j},$$

[4] A folklore example with complete proofs can be found at `http://cstheory.stackexchange.com/questions/14471/reverse-chernoff-bound`.

we get

$$\begin{aligned}
&\Pr\left(\mathcal{E}^{(s)}_{\bar{m}>} \mid \mathbf{C}^{(s-1)} = \hat{\mathbf{c}}\right) \\
&\leqslant \Pr\left(C^{(s)}_{\bar{m}} > (1-\delta)\mathbb{E}\left[C^{(s)}_{\bar{m}} \mid \mathbf{C}^{(s-1)} = \hat{\mathbf{c}}\right] \mid \mathbf{C}^{(s-1)} = \hat{\mathbf{c}}\right) \\
&\leqslant 1 - e^{-4\beta j^2 \log n}. \qquad (82)
\end{aligned}$$

By the law of total probability, we can thus saturate with respect to all $\hat{\mathbf{c}} \in \mathcal{S}_{\mathrm{m}}$ and from (82) we obtain that

$$\Pr\left(\mathcal{E}^{(s)}_{\mathrm{m}>} \,\middle|\, \mathcal{E}^{(s-1)}_{\mathrm{m}>} \wedge (\mathbf{C}^{(0)} = \mathbf{c})\right) \leqslant 1 - e^{-4\beta j^2 \log n},$$

which proves (81).

On the other hand, from Chernoff bounds (Lemma 76) and the fact that

$$\mathbb{E}\left[C^{(s)}_{\bar{m}} \mid \mathbf{C}^{(s-1)} = \hat{\mathbf{c}}\right] \leqslant \frac{n}{j},$$

it follows that

$$\Pr\left(C^{(s)}_{\bar{m}} < \left\lfloor \frac{n}{j} \right\rfloor - \alpha\sqrt{jn\log n} \,\middle|\, \bigwedge_{r=1}^{s-1} \mathcal{E}^{(r)}_{\mathrm{m}>} \wedge \mathbf{C}^{(0)} = \mathbf{c}\right) \leqslant e^{-\frac{\alpha^2}{6} j^2 \log n}. \qquad (83)$$

Finally, substituting (82) and (83) into (80), the result follows by choosing α large enough in (83). gg

We now provide detailed proofs of the two technical lemmas of the dropping state.

LEMMA 29 (Dropping Stage 1). *Let $\mathbf{c}$ be any configuration with $j \leqslant n^{1/3-\varepsilon}$ active opinions, where $\varepsilon > 0$ is an arbitrarily-small positive constant, and such that an opinion i exists with $c_i \leqslant n/j - \sqrt{jn\log n}$. Within $t = \mathcal{O}(j\log n)$ rounds opinion i becomes $\mathcal{O}\left(j^2 \log n\right)$, w.h.p.*

PROOF. We first prove that the decreasing rate of C_i depends on its value at the end of the previous round. More formally, if we are in a configuration satisfying the hypotheses of the lemma, we have

$$\begin{aligned}
&\Pr\left(C_i^{(t)} > c_i^{(t-1)}\left(1 - \frac{1}{2}\left(\frac{1}{j} - \frac{c_i^{(t-1)}}{n}\right)\right)\right) \\
&= \Pr\left(C_i^{(t)} > c_i^{(t-1)}\left(1 - \left(\frac{1}{j} - \frac{c_i^{(t-1)}}{n}\right)\right)(1+\delta)\right),
\end{aligned}$$

where

$$\delta = \frac{\frac{1}{2}\left(\frac{1}{j} - \frac{c_i^{(t-1)}}{n}\right)}{1 - \left(\frac{1}{j} - \frac{c_i^{(t-1)}}{n}\right)}.$$

Using Lemma 26 and applying Chernoff bound (Lemma 76) we thus get

$$\begin{aligned}
&\Pr\left(C_i^{(t)} > c_i^{(t-1)}\left(1-\frac{1}{2}\left(\frac{1}{j}-\frac{c_i^{(t-1)}}{n}\right)\right)\right)\\
&\leqslant \exp\left(-\frac{\delta^2}{3}\left(1-\left(\frac{1}{j}-\frac{c_i^{(t-1)}}{n}\right)\right)c_i^{(t-1)}\right)\\
&\overset{(a)}{=} \exp\left(-\frac{\delta}{3}\left(\frac{1}{2}\left(\frac{1}{j}-\frac{c_i^{(t-1)}}{n}\right)\right)c_i^{(t-1)}\right)\\
&\overset{(b)}{<} \exp\left(-\frac{1}{3}\left(\frac{1}{2}\left(\frac{1}{j}-\frac{c_i^{(t-1)}}{n}\right)\right)^2 c_i^{(t-1)}\right) \overset{(c)}{=} n^{-\Theta(1)},
\end{aligned} \tag{84}$$

where

- (a) follows from the definition of δ,
- (b) follows by (upper) bounding the denominator of δ by 1, which is always possible since $c_i/n - 1/j < 0$ by hypothesis, and
- (c) follows from the fact that $c_i \geqslant j^2 \log n$ and that the function $x\,(1-x)^2$ is decreasing iff $x \in (1/3, 1)$, with $x = jc_i/n$.

Finally, we can iteratively apply (84) as long as we have at most j active opinions and $C_i^{(t)}$ is bigger than $j^2 \log n$: By standard concentration arguments we get that the time to reach this threshold is $\mathcal{O}(j \log n)$, w.h.p. ⊠

LEMMA 30 (Dropping Stage 2). *Let* $\mathbf{c}$ *be any configuration with* $j \leqslant n^{1/3-\varepsilon}$ *active opinions, where* $\varepsilon > 0$ *is an arbitrarily-small positive constant, and such that an opinion* i *exists with* $c_i \leqslant n/(2j)$. *Within* $t = \mathcal{O}(j \log n)$ *rounds opinion* i *disappears with probability at least* $1/2$.

PROOF. Let J be the set of active opinions. By conditioning on all the configurations $\hat{\mathbf{c}} = (\hat{c}_\ell : \ell \in J)$ that the system can take at round $t-1$, we can bound the expectation of $C_i^{(t)}$ as follows

$$\begin{aligned}
&\mathbb{E}\left[C_i^{(t)} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right]\\
&= \sum_{\hat{\mathbf{c}}} \mathbb{E}\left[C_i^{(t)} \,|\, \mathbf{C}^{(t-1)} = \hat{\mathbf{c}}\right] \Pr\left(\mathbf{C}^{(t-1)} = \hat{\mathbf{c}} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right)\\
&\leqslant \left(1-\frac{1}{2j}\right) \sum_{\hat{\mathbf{c}}\,:\,\hat{c}_i \leqslant n/(2j)} \hat{c}_i \cdot \Pr\left(\mathbf{C}^{(t-1)} = \hat{\mathbf{c}} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right)\\
&\qquad + n \cdot \sum_{\hat{\mathbf{c}}\,:\,\hat{c}_i > n/(2j)} \Pr\left(\mathbf{C}^{(t-1)} = \hat{\mathbf{c}} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right)\\
&\leqslant \left(1-\frac{1}{2j}\right) \mathbb{E}\left[C_i^{(t-1)} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right]
\end{aligned}$$

$$+ n \cdot \Pr\left(C_i^{(t-1)} > \frac{n}{2j} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right), \tag{85}$$

where we used that, for any configuration $\hat{\mathbf{c}}$ with $\hat{c}_i \leqslant n/(2j)$, Lemma 26 gives the bound

$$\mathbb{E}\left[C_i^{(t)} \,|\, \mathbf{C}^{(t-1)} = \hat{\mathbf{c}}\right] \leqslant \hat{c}_i \left(1 - \frac{1}{2j}\right).$$

Moreover, if $j \leqslant n^{1/3-\varepsilon}$, from the Chernoff bound (Lemma 76) it follows that

$$\Pr\left(C_i^{(t)} > \frac{n}{2j} \,|\, \mathbf{C}^{(t-1)} = \hat{\mathbf{c}}\right) \leqslant e^{-\Theta(n^{\varepsilon})} \tag{86}$$

for any such configuration $\hat{\mathbf{c}}$. Hence, for any t we can bound the second term in (85) as follows:

$$\begin{aligned}
&\Pr\left(C_i^{(t)} > \frac{n}{2j} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right) \\
&\leqslant \Pr\left(\exists \bar{t} = 1, \dots, t \,:\, \left(C_i^{(\bar{t})} > \frac{n}{2j}\right) \wedge \left(C_i^{(\bar{t}-1)} \leqslant \frac{n}{2j}\right) \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right) \\
&\leqslant \sum_{\bar{t}=1}^{t} \Pr\left(\left(C_i^{(\bar{t})} > \frac{n}{2j}\right) \wedge \left(C_i^{(\bar{t}-1)} \leqslant \frac{n}{2j}\right) \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right) \\
&= \sum_{\bar{t}=1}^{t} \sum_{\hat{\mathbf{c}} \,:\, \hat{c}_i \leqslant \frac{n}{2j}} \Pr\left(C_i^{(\bar{t})} > \frac{n}{2j} \,|\, \mathbf{C}^{(\bar{t}-1)} = \hat{\mathbf{c}}\right) \\
&\qquad \cdot \Pr\left(\mathbf{C}^{(\bar{t}-1)} = \hat{\mathbf{c}} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right) \leqslant t e^{-\Theta(n^{\varepsilon})}
\end{aligned}$$

where in the last inequality we used (86). Thus for any $t = \mathrm{poly}(n)$ the following recursive relation holds

$$\mathbb{E}\left[C_i^{(t)} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right] \leqslant \left(1 - \frac{1}{2j}\right) \mathbb{E}\left[C_i^{(t-1)} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right] + \exp\left(-n^{\frac{\varepsilon}{2}}\right),$$

that is

$$\mathbb{E}\left[C_i^{(t)} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right] \leqslant \left(1 - \frac{1}{2j}\right)^t \frac{n}{2j} + e^{-n^{\varepsilon/3}}. \tag{87}$$

From (87), for $t = 2j(\log n + 1)$ we get $\mathbb{E}\left[C_i^{(t)} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right] \leqslant 1/2$ and since $C_i^{(t)}$ takes non-negative integer values, the thesis follows from Markov's inequality. gs

LEMMA 36 (Bounded Jump). *Let $\tilde{\mathbf{c}}$ be any configuration such that $|B| = j$ and $\sum_{i \in \bar{\mathcal{C}}} \tilde{c}_i \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$. Consider the stochastic process $\{\tilde{Y}_t\}_t$ defined as*

$\tilde{Y}_t = \left\lfloor \frac{n}{j} \right\rfloor - \tilde{C}_{\mathrm{m}}^{(t)}$ *and define the stopping time*

$$\tau = \inf\{t \in \mathbb{N} : \tilde{Y}_t \geqslant \sqrt{jn \log n} \vee \big(\sum_{i \in \bar{\mathcal{C}}} \tilde{C}_i \geqslant \gamma\sqrt{n}/k^{\frac{3}{2}}\big) \vee (S^{(t-1)} \not\subseteq S^{(t)})\}.$$

It holds that

$$\Pr\left(\tilde{Y}_\tau > \alpha\sqrt{jn \log n} \,|\, \mathbf{C}^{(0)} = \mathbf{c}\right) \leqslant \frac{1}{n}.$$

Sketch of Proof. The proof of this Lemma follows from minor modifications of the proof of Lemma 35. In particular, the argument is based on the following observations:

1. The event defining the stopping time τ is in this case

$$\mathcal{E}^{(t)} = \big(\tilde{Y}_t \geqslant (\sqrt{jn \log n}) \vee \big(\sum_{i \in \bar{\mathcal{C}}} \tilde{C}_i \geqslant \gamma\sqrt{n}/k^{\frac{3}{2}}\big) \vee (S^{(t-1)} \not\subseteq S^{(t)})\big).$$

The negated of this event is

$$\neg\mathcal{E}^{(t)} = \big(\tilde{Y}_t \leqslant (\sqrt{jn \log n}) \wedge \big(\sum_{i \in \bar{\mathcal{C}}} \tilde{C}_i \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}\big) \wedge (S^{(t-1)} \subseteq S^{(t)})\big),$$

which implies the event "$\tilde{Y}_t \leqslant \sqrt{jn \log n}$".

2. Proceeding like in the proof of Lemma 35, we can write an expression that is similar to (80), with the generic conditioning event

$$C_{\mathrm{m}}^{(s)} > \left\lfloor \frac{n}{j} \right\rfloor - \sqrt{jn \log n},$$

replaced by $\neg\mathcal{E}^{(s)}$. The conditioned event

$$C_\ell^{(t)} < \left\lfloor \frac{n}{j} \right\rfloor - \alpha\sqrt{jn \log n},$$

is instead replaced by the event

$$\left(C_\ell^{(t)} < \left\lfloor \frac{n}{j} \right\rfloor - \alpha\sqrt{jn \log n}\right) \wedge \mathcal{E}^{(t)}.$$

Now, note that the event

$$C_\ell^{(t)} < \left\lfloor \frac{n}{j} \right\rfloor - \alpha\sqrt{jn \log n},$$

again implies $\mathcal{E}^{(t)}$. Hence, we can still write (80), from which the proof requires minor adaptations w.r.t. Lemma 35. ▣

Since the adversary, at round t, may decide what to do based on the full history of the process up to time t, the stochastic process $\{\tilde{\mathbf{C}}^{(t)}\}_t$ may not be a Markov process anymore. Thus, we need a more general version of Lemma 27.

LEMMA 37. *Let $\{X_t\}_t$ be a discrete time stochastic process with a finite state space Ω, let $f_t : \Omega^t \to N$ be a function mapping histories of the process in non-negative integer numbers, and let $\{Y_t\}_t$ be the stochastic process over N defined by $Y_t = f_t(X_0, \dots, X_t)$. Let $m \in N$ be a "target value", let $A \subseteq \Omega$ be an arbirary subset of states, and let*

$$\tau = \inf\{t \in \mathbb{N} : Y_t \geqslant m \text{ or } X_t \notin A\}$$

be the random variable indicating the first time X_t exits from set A or Y_t reaches or exceeds value m. Assume that, for every sequence of states $x_0, \dots, x_t \in A$ with $f_t(x_0, \dots, x_t) \leqslant m - 1$, it holds that

(1) (Positive drift). For some $\lambda > 0$, it holds

$$\mathbb{E}\left[Y_{t+1} \mid X_0 = x_0, \dots, X_t = x_t\right] \geqslant f_t(x_0, \dots, x_t) + \lambda,$$

(2) (Bounded jumps). For some $\alpha > 1$

$$\Pr\left(Y_{\tau \mid X_t = x} \geqslant \alpha m\right) \leqslant \alpha m/n.$$

Then, for every starting state $x \in A$, it holds that

$$\mathbb{E}\left[\tau \mid X_t = x\right] \leqslant 2\alpha \frac{m}{\lambda}.$$

PROOF. The proof is a straight adaptation of the proof of Lemma 27, in which we take into account the full history of the process.

Consider the stochastic process $Z_t = Y_t - \lambda t$. For any sequence of states $x_0, \dots, x_t \in A$ with $f_t(x_0, \dots, x_t) \leqslant m - 1$ it holds that

$$\begin{aligned}
&\mathbb{E}\left[Z_{t+1} \mid X_0 = x_0, \dots, X_t = x_t\right] \\
&= \mathbb{E}\left[Y_{t+1} \mid X_0 = x_0, \dots, X_t = x_t\right] - \lambda(t+1) \\
&\geqslant f_t(x_0, \dots, x_t) + \lambda - \lambda(t+1) \\
&\geqslant f_t(x_0, \dots, x_t) - \lambda t,
\end{aligned}$$

where in the inequality we used Hypothesis *1*. Thus, Z_t is a *submartingale* up to the stopping time τ. Moreover, since $|Y_t| \leqslant n$ then

$$|Z_{t+1} - Z_t| \leqslant n + \lambda$$

and, together with Hypothesis *1* this implies $\mathbb{E}\left[\tau \mid X_t = x\right] < \infty$. Thus, we can apply *Doob's Optional Stopping Theorem* (Theorem 27, pag. 272). It follows that

$$\mathbb{E}\left[Z_\tau \mid X_t = x\right] \geqslant \mathbb{E}\left[Z_0 \mid X_t = x\right] = f_0(x),$$

and since

$$\mathbb{E}\left[Z_\tau \mid X_t = x\right] = \mathbb{E}\left[Y_\tau \mid X_t = x\right] - \lambda \mathbb{E}\left[\tau \mid X_t = x\right],$$

we have that

$$\mathbb{E}\left[\tau \mid X_t = x\right] \leqslant \frac{\mathbb{E}_x\left[Y_\tau\right] - f_0(x)}{\lambda} \leqslant \frac{\mathbb{E}\left[Y_\tau \mid X_t = x\right]}{\lambda}.$$

Finally, we get

$$\mathbb{E}\left[Y_\tau \,|\, X_0 = 0\right] = \sum_{j=1}^{\lfloor \alpha m \rfloor} j \Pr\left(Y_\tau = j \,|\, X_0 = 0\right) + \sum_{j=\lfloor \alpha m \rfloor + 1}^{n} j \Pr\left(Y_\tau = j \,|\, X_0 = 0\right)$$
$$\leqslant (\alpha m) + n \Pr\left(Y_\tau > \alpha m \,|\, X_0 = 0\right) \leqslant 2\,(\alpha m)\,,$$

where in the last inequality we used Hypothesis *2*. ⊠

With the following lemma, we generalize Lemma 28 to the adversarial setting.

LEMMA 31 (Symmetry-Breaking Stage with Adversary). *Let $\tilde{\mathbf{c}}$ be any configuration such that $|B| = j$ and $\sum_{i\in\bar{\mathcal{C}}} \tilde{c}_i \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$. Within $t = \mathcal{O}(j^2 \log^{1/2} n)$ rounds, with probability at least $1/2$ it holds that*

i) $|B| = j$, $\sum_{i\in\bar{\mathcal{C}}} \tilde{C}_i \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$, and
ii) there exists an $i \in B^{(t)}$ such that $\tilde{C}_i^{(t)} \leqslant n/j - \sqrt{jn\log n}$.

PROOF. We proceed by adapting the proof of Lemma 28. Let $\tilde{\mathbf{C}}^{(0)} = \tilde{\mathbf{c}}$ be the initial configuration. Let us consider the stochastic process $\{\tilde{Y}_t\}_{t\geqslant 0}$ defined as

$$\tilde{Y}_t = \left\lfloor \frac{n}{j} \right\rfloor - \tilde{C}_{\mathrm{m}}^{(t)} \quad \text{where} \quad \tilde{C}_{\mathrm{m}}^{(t)} = \min\{\tilde{C}_i^{(t)} : i \in B^{(t)}\}.$$

We are interested in estimating the expected value of

$$\tau = \inf\{t \in \mathbb{N} \,:\, \tilde{Y}_t \geqslant (\sqrt{jn\log n}) \vee \big(\sum_{i\in\bar{\mathcal{C}}} \tilde{C}_i \geqslant \gamma\sqrt{n}/k^{\frac{3}{2}}\big)$$
$$\vee\ (S^{(t-1)} \not\subseteq S^{(t)})\}.$$

Now we show that $\{\tilde{Y}_t\}_t$ satisfies the Hypotheses *1* and *2* of Lemma 27 with

$$A = \big(\sum_{i\in\bar{\mathcal{C}}} \tilde{C}_i \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}\big) \vee (S^{(t-1)} \subseteq S^{(t)})$$

and $\lambda = \varepsilon\sqrt{n}/j^{3/2}$, for a suitable constant $\varepsilon > \alpha$.

1. Let $\tilde{\mathbf{c}}$ be any configuration such that $\tilde{c}_{\mathrm{m}} > n/j - \sqrt{jn\log n}$. Now we prove that

$$\mathbb{E}\left[\tilde{C}_{\mathrm{m}}^{(t+1)} \,|\, \tilde{\mathbf{C}}^{(t)} = \tilde{\mathbf{c}}\right] \leqslant \tilde{c}_{\mathrm{m}} - \varepsilon\frac{\sqrt{n}}{j^{3/2}}. \tag{88}$$

Case $\tilde{c}_{\mathrm{m}} > n/j - 2\varepsilon\sqrt{n/j}$. Observe that, in this case, random variables $\{C_i^{t+1} : i \in B\}$ have standard deviation is $\Omega(\sqrt{n/j})$. Moreover they are binomial and negatively associated. Hence, by choosing ε small enough, from the Central Limit Theorem we have that

$$\Pr\left(\exists\, i \in B \text{ such that } C_i^{(t+1)} \leqslant \frac{n}{j} - 6\varepsilon \cdot \sqrt{\frac{n}{j}}\right) \geqslant 1/2.$$

We thus get

$$\begin{aligned}\mathbb{E}\left[\tilde{C}_{\mathtt{m}}^{(t+1)} \,|\, \tilde{\mathbf{C}}^{(t)} = \tilde{\mathbf{c}}\right] &\leqslant \frac{1}{2}\left(\frac{n}{j} - 6\varepsilon \cdot \sqrt{\frac{n}{j}}\right) + \frac{1}{2}\cdot\frac{n}{j} + \frac{\beta\sqrt{n}}{k^{\frac{5}{2}}\log n} \\ &= \frac{n}{j} - 2\varepsilon\sqrt{\frac{n}{j}} + \frac{\beta\sqrt{n}}{k^{\frac{5}{2}}\log n} \\ &\leqslant \tilde{c}_{\mathtt{m}} - \varepsilon\sqrt{\frac{n}{j}} \leqslant \tilde{c}_{\mathtt{m}} - \varepsilon\frac{\sqrt{n}}{j^{3/2}}.\end{aligned}$$

Case $\tilde{c}_{\mathtt{m}} \leqslant n/j - 2\varepsilon\sqrt{n/j}$. (88) easily follows from Lemma 32. Indeed, let $i \in B$ be an opinion such that $\hat{c}_i = \hat{c}_{\mathtt{m}}$, then

$$\begin{aligned}\mathbb{E}\left[\tilde{C}_{\mathtt{m}}^{(t+1)} \,|\, \tilde{\mathbf{C}}^{(t)} = \tilde{\mathbf{c}}\right] &\leqslant \mathbb{E}\left[\tilde{C}_{i}^{(t+1)} \,|\, \tilde{\mathbf{C}}^{(t)} = \tilde{\mathbf{c}}\right] \\ &\leqslant \tilde{c}_i\left(1 + \frac{\tilde{c}_i + \alpha\sqrt{n/k}}{n} - \frac{1}{j}\right) \\ &\leqslant \tilde{c}_i\left(1 - \frac{2\varepsilon}{\sqrt{nj}} + \frac{\alpha}{\sqrt{kn}}\right) \\ &\leqslant \tilde{c}_i - \frac{\varepsilon\sqrt{n}}{j^{3/2}} = \tilde{c}_{\mathtt{m}} - \varepsilon\frac{\sqrt{n}}{j^{3/2}},\end{aligned}$$

where we used the case's condition and $\tilde{c}_i = \tilde{c}_{\mathtt{m}} \geqslant n/(2j)$.

2. Since random variables $\{\tilde{C}_i^{(t)} : i \in B^{(t)}\}$ are binomial conditional on the configuration at round $t-1$, from the Chernoff bound (Lemma 76) it follows that

$$\Pr\left(\tilde{Y}_\tau \geqslant \alpha\sqrt{jn\log n} \,|\, \tilde{\mathbf{C}}^{(0)} = \tilde{\mathbf{c}}\right) \leqslant \frac{1}{n}, \quad \text{for some constant} \quad \alpha > 1. \tag{89}$$

See Lemma 36 for the formal statement of the last fact.

From (88) and (89) we have that $\{\tilde{Y}_t\}_t$ satisfies the hypotheses of Lemma 37 with $m = \sqrt{jn\log n}$, $\lambda = \varepsilon\sqrt{n}/j^{3/2}$ and

$$A = \big(\sum_{i\in\bar{\mathcal{C}}} \tilde{C}_i \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}\big) \vee (S^{(t-1)} \subseteq S^{(t)}).$$

Moreover, by iteratively applying Lemma 33, we have that, for any $t = \mathcal{O}(n^2)$, it holds w.h.p.

$$\big(\sum_{i\in\bar{\mathcal{C}}} \tilde{C}_i^{(t)} \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}\big) \vee (S^{(t-1)} \subseteq S^{(t)}).$$

Thus, from Markov's inequality, for $t = 2j^2\sqrt{\log n}$, we have that

$$\begin{aligned}&\Pr\big(\forall i \in B : \big(C_i^{(t)} \leqslant n/j - \sqrt{jn\log n}\big) \wedge \big(\sum_{i\in\bar{\mathcal{C}}} \tilde{C}_i^{(t)} \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}\big) \\ &\qquad \wedge (S^{(0)} \subseteq S^{(t)}) \,|\, \tilde{\mathbf{C}} = \tilde{\mathbf{c}}\big) \\ &\geqslant \Pr\left(\hat{\tau} \leqslant 2j^2\sqrt{\log n} \,|\, \mathbf{C}^{(0)} = \tilde{\mathbf{c}}\right) \geqslant \frac{1}{3}\end{aligned}$$

where

$$\hat{\tau} = \inf\{t \in \mathbb{N} : \tilde{Y}_t \geqslant \sqrt{jn \log n}\}.$$

gg

In the next lemmas, we provide the analogous versions of lemmas 29 and 30 in the adversarial setting.

LEMMA 32 (Dropping Stage 1 with Adversary). *Let $\tilde{\mathbf{c}}$ be any configuration such that $|B| \leqslant j$ and $\sum_{i \in \bar{\mathcal{C}}} \tilde{c}_i^{(t)} \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$. For some constant $\alpha > 0$, for any opinion i such that $\tilde{c}_i \geqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$, it holds*

$$\begin{aligned}
&\mathbb{E}\left[C_i^{(t+1)} \,\middle|\, \tilde{\mathbf{C}}^{(t)} = \tilde{\mathbf{c}}\right] \leqslant \tilde{c}_i\left(1 - \frac{1}{j} + \frac{\tilde{c}_i + \alpha\sqrt{n/k}}{n}\right) \\
&\mathbb{E}\left[\tilde{C}_i^{(t+1)} \,\middle|\, \tilde{\mathbf{C}}^{(t)} = \tilde{\mathbf{c}}\right] \leqslant \tilde{c}_i(1 - \eta(i,j)), \text{ where} \qquad (78) \\
&\eta(i,j) = \min\left\{\frac{1}{j} - \frac{\tilde{c}_i + \alpha\sqrt{n/k}}{n}\ ,\ \frac{1}{2}\left(\frac{1}{j} - \frac{\tilde{c}_i}{n}\right)\right\}
\end{aligned}$$

PROOF. Similarly to the proof of Lemma 26 we have

$$\begin{aligned}
&\mathbb{E}\left[C_i^{(t+1)} \,\middle|\, \tilde{\mathbf{C}}^{(t)} = \tilde{\mathbf{c}}\right] \\
&\leqslant \tilde{c}_i\left(1 + \frac{\tilde{c}_i}{n} - \frac{\sum_j \tilde{c}_j^2}{n^2}\right) \leqslant \tilde{c}_i\left(1 + \frac{\tilde{c}_i}{n} - \frac{\sum_{j\in B} \tilde{c}_j^2}{n^2}\right) \\
&\leqslant \tilde{c}_i\left(1 + \frac{\tilde{c}_i}{n} - \frac{\sum_{j\in B}\left(\frac{n-(k-j+1)\gamma\sqrt{n}/k^{\frac{3}{2}}}{j}\right)^2}{n^2}\right) \\
&\leqslant \tilde{c}_i\left(1 + \frac{\tilde{c}_i}{n} - \frac{\sum_{j\in B}(n - \alpha/4\sqrt{n/k})^2}{j^2n^2}\right) \\
&\leqslant \tilde{c}_i\left(1 + \frac{\tilde{c}_i}{n} - \frac{1}{j} + \frac{\alpha/2\sqrt{n/k}}{jn}\right) \\
&\leqslant \tilde{c}_i\left(1 - \frac{n/j - \tilde{c}_i - \alpha/2\sqrt{n/k}}{n}\right).
\end{aligned}$$

Taking into account any possible action of the adversary, we thus get that

$$\begin{aligned}
&\mathbb{E}\left[\tilde{C}_i^{(t+1)} \,\middle|\, \tilde{\mathbf{C}}^{(t)} = \tilde{\mathbf{c}}\right] \\
&= \mathbb{E}\left[C_i^{(t+1)} \,\middle|\, \tilde{\mathbf{C}}^{(t)} = \tilde{\mathbf{c}}\right] + \mathbb{E}\left[D_i^{(t+1)} \,\middle|\, \tilde{\mathbf{C}}^{(t)} = \tilde{\mathbf{c}}\right] \\
&\leqslant \tilde{c}_i \left(1 - \frac{n/j - \tilde{c}_i - \alpha/2\sqrt{n/k}}{n}\right) + F \\
&\leqslant \tilde{c}_i \left(1 - \frac{n/j - \tilde{c}_i}{n} + \frac{2\max\left\{\alpha/2\sqrt{n/k}, Fn/\tilde{c}_i\right\}}{n}\right). \qquad (90)
\end{aligned}$$

By distinguishing the cases $\tilde{c}_i \geqslant n/(3j)$ or $\tilde{c}_i < n/(3j)$, from (90) we get (78). ⊠

LEMMA 34 (Dropping Stage 2 with Adversary). *Assume that, at round t', $\tilde{\mathbf{c}}^{(t')}$ is such that*

- $\sum_{i\in\bar{\mathcal{C}}} c_i \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$,
- $|B^{(t')}| = j$, *and*
- *an $i \in B^{(t')}$ exists such that*

$$\gamma\sqrt{n}/k^{\frac{3}{2}} \leqslant c_i^{(t')} \leqslant n/j - \sqrt{kn\log n}.$$

Then, a round $t'' = t' + O(k\log n)$ exists such that w.h.p.

- $\sum_{i\in\bar{\mathcal{C}}} \tilde{C}_i^{(t'')} \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$,
- $i \in S^{(t'')}$ *and*
- $|B^{(t'')}| \leqslant j-1$.

PROOF. By iteratively applying Lemma 33, we have that for each $t \in \{t', \ldots, t''-1\}$ it holds $\sum_{i\in\bar{\mathcal{C}}} \tilde{C}_i^{(t)} \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$ and $i \in S^{(t)}$, w.h.p.

To prove that $|B^{(t'')}| \leqslant j-1$, we first prove that, for each round $t \in \{t'+1, \ldots, t''\}$, $\tilde{C}_i^{(t)}$ decreases by a certain extent that depends on $\tilde{c}_i^{(t-1)}$, w.h.p., regardless of what the adversary does.

Let

$$\psi = \left(\frac{1}{j} - \frac{\tilde{c}_i^{(t-1)} + \alpha\sqrt{\frac{n}{k}}}{n}\right).$$

If we are in a configuration satisfying the hypotheses of the lemma, we have

$$\Pr\left(C_i^{(t)} > \tilde{c}_i^{(t-1)}\left(1 - \frac{\psi}{2}\right)\right) = \Pr\left(C_i^{(t)} > \tilde{c}_i^{(t-1)}\left(1 - \psi(1+\delta)\right)\right),$$

where

$$\delta = \frac{\psi}{2(1-\psi)}.$$

Thus, using Lemma 32 and applying the Chernoff bound (Lemma 76) we have

$$\begin{aligned}\Pr\left(C_i^{(t)} > \tilde{c}_i^{(t-1)}\left(1-\frac{\psi}{2}\right)\right) &\leqslant \exp\left(-\frac{\delta^2}{3}\psi\tilde{c}_i^{(t-1)}\right)\\ &\overset{(a)}{<} \exp\left(-\frac{1}{3}\left(\frac{1}{2}\psi\right)^2\tilde{c}_i^{(t-1)}\right) \overset{(b)}{=} n^{-\Theta(1)},\end{aligned} \tag{91}$$

where (a) follows from the definition of δ and the fact that its denominator is smaller than 1, and (b) follows by minimizing $\psi^2\tilde{c}_i^{(t-1)}$ for

$$\gamma\sqrt{n}/k^{\frac{3}{2}} \leqslant c_i^{(t')} \leqslant n/j - \sqrt{kn\log n}.$$

It follows that w.h.p.

$$\tilde{C}_i^{(t)} = C_i^{(t)} + D_i^{(t)} \leqslant \tilde{c}_i^{(t-1)}\left(1-\frac{\psi}{2}\right) + F \leqslant \tilde{c}_i^{(t-1)} \tag{92}$$

Thus, we can iteratively apply (92) until $\tilde{c}_i^{(t-1)} \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$, w.h.p. We next prove that this happens within $\mathcal{O}(k\log n)$ rounds, w.h.p. Interestingly, showing that, within $\mathcal{O}(k\log n)$ rounds, C_i decreases to a costant fraction of its value at the beginning of the dropping stage does not seem obvious. For this reason, we consider the evolution of the displacement $\frac{n}{j} - C_i$, which seems analytically more tractable. To this purpose, note that (91) implies that w.h.p.

$$\begin{aligned}\frac{n}{j} - C_i^{(t)} &\geqslant \frac{n}{j} - c_i^{(t-1)} + \frac{c_i^{(t-1)}}{2}\left(\frac{1}{j} - \frac{c_i^{(t-1)} + \alpha\sqrt{n/k}}{n}\right)\\ &\overset{(a)}{=} \frac{n}{j} - c_i^{(t-1)} + \frac{c_i^{(t-1)}}{2}\left(\frac{1}{j} - \frac{c_i^{(t-1)}}{n}\right)\left(1 - \frac{\alpha\sqrt{n/k}}{\frac{1}{j} - \frac{c_i^{(t-1)}}{n}}\right)\\ &= \frac{n}{j} - c_i^{(t-1)} + \frac{c_i^{(t-1)}}{2}\left(\frac{1}{j} - \frac{c_i^{(t-1)}}{n}\left(1 + \frac{\alpha}{\log n}\right)\right)\\ &= \left(\frac{n}{j} - c_i^{(t-1)}\right)\left(1 + \alpha_1\frac{c_i^{(t-1)}}{2n}\right),\end{aligned} \tag{93}$$

for some constant $\alpha_1 > 0$, where in (a) we have used that $n/j - c_i^{(t-1)} \geqslant \sqrt{kn\log n}$.

We can now conclude the proof of Lemma 34. We first prove that $C_i \leqslant n/(2j)$ within $\mathcal{O}(k\log n)$ steps, w.h.p. To this purpose note that, from the hypotheses, at the beginning of the dropping stage it holds

$$\frac{n}{j} - c_i \geqslant \sqrt{kn\log n}.$$

Furthermore, for some positive constants α_2 and α_3, as long as $C_i \geqslant \alpha_3 n/j$, it holds

$$1 + \alpha_1 \frac{c_i}{n} \geqslant 1 + \frac{\alpha_2}{j}.$$

Hence, after $\mathcal{O}(k \log n)$ steps, we have w.h.p.

$$\frac{n}{j} - c_i \geqslant (1 - \alpha_3)\frac{n}{j},$$

which in turn implies $c_i \leqslant \alpha_3 n/j$. Once $c_i \leqslant \alpha_3 n/j$, using again (93) we have that C_i decreases by a factor $1 - \Omega(1/j)$ in every round, w.h.p. By standard concentration arguments we obtain that eventually $c_i \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$ within $\mathcal{O}(k \log n)$ more steps, w.h.p. gg

Finally, it remains to prove 33.

LEMMA 33 (Small Stays Small). *If $\tilde{\mathbf{c}}^{(t)}$ is such that $\sum_{i\in\bar{\mathcal{C}}} \tilde{c}_i^{(t)} \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$, then $\sum_{i\in\bar{\mathcal{C}}} \tilde{C}_i^{(t+1)} \leqslant \gamma\sqrt{n}/k^{\frac{3}{2}}$ and $S^{(t)} \subseteq S^{(t+1)}$, w.h.p.*

PROOF. From Lemma 32, for each $i \in S^{(t)}$ we have that

$$\mathbb{E}\left[C_i^{(t+1)} \,\middle|\, \tilde{\mathbf{C}}^{(t)} = \tilde{\mathbf{c}}\right] \leqslant \tilde{c}_i \left(1 + \frac{\tilde{c}_i}{n} - \frac{1}{k}\right).$$

From a direct application of the Chernoff bound (Lemma 76) to $C_i^{(t+1)}$, and taking into account any possible action of the adversary, we thus get that w.h.p.

$$\tilde{C}_i^{(t+1)} = C_i^{(t+1)} + D_i^{(t+1)} \leqslant \gamma\frac{\sqrt{n}}{k^{\frac{3}{2}}}\left(1 - \frac{1}{4k}\right) + F \leqslant \gamma\frac{\sqrt{n}}{k^{\frac{3}{2}}},$$

that is, $i \in S^{(t)}$, w.h.p. Analogously, we have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i\in\bar{\mathcal{C}}^{(t)}} C_i^{(t+1)} \,\middle|\, \tilde{\mathbf{C}}^{(t)} = \tilde{\mathbf{c}}\right] \\
&\leqslant \sum_{i\in\bar{\mathcal{C}}} \tilde{c}_i^{(t)} \left(1 + \frac{\tilde{c}_i}{n} - \frac{1}{k}\right) \leqslant \gamma\frac{\sqrt{n}}{k^{\frac{3}{2}}}\left(1 - \frac{1}{2k}\right),
\end{aligned}$$

and then, by applying the Chernoff bound (Lemma 76), we get that w.h.p.

$$\begin{aligned}
\sum_{i\in\bar{\mathcal{C}}^{(t)}} \tilde{C}_i^{(t+1)} &= \sum_{i\in\bar{\mathcal{C}}^{(t)}} C_i^{(t+1)} + \sum_i D_i^{(t+1)} \\
&\leqslant \gamma\frac{\sqrt{n}}{k^{\frac{3}{2}}}\left(1 - \frac{1}{4k}\right) + F \leqslant \gamma\frac{\sqrt{n}}{k^{\frac{3}{2}}},
\end{aligned}$$

concluding the proof. gg

CHAPTER 6

Undecided-State Dynamics

In this chapter we prove the results presented in Section 2.3, continuing the investigation of Chapter 5 about efficient dynamics for the *plurality consensus* problem in the $\mathcal{PULL}$ model over a network of n anonymous agents.

We consider the *Undecided-State* dynamics, a well-known protocol which uses just one more state (the *undecided* one) than those necessary to store opinions. We show that the speed of convergence of this protocol depends on the initial opinion configuration as a whole, not just on the gap between the plurality and the second largest opinion community. This dependence is best captured by the notion of *monochromatic distance* $\mathsf{md}(\mathbf{c})$, which measures the distance of the initial opinion configuration $\mathbf{c}$ from the closest monochromatic one. In the complete graph, we prove that, for a wide range of the input parameters, this dynamics converges within $O(\mathsf{md}(\mathbf{c}) \log n)$ rounds. We prove that this upper bound is almost tight in the strong sense: Starting from *any* opinion configuration $\mathbf{c}$, the convergence time is $\Omega(\mathsf{md}(\mathbf{c}))$.

Finally, we adapt the Undecided-State dynamics to obtain a fast, random walk-based protocol for plurality consensus on *regular expanders*. This protocol converges in $O(\mathsf{md}(\mathbf{c}) \operatorname{polylog}(n))$ rounds using only $\operatorname{polylog}(n)$ local memory. A key-ingredient to achieve the above bounds is the analysis of the maximum node congestion that results from performing n parallel random walks on regular expanders.

6.1. Warm Up Before the Analysis

Recall that in the plurality consensus problem each of the n (anonymous) agents in the system supports an initial opinion or *color* out of a set of $k = k(n) \in [n]$ possible ones. At the onset, the number of agents supporting the *plurality* opinion $j \in [k]$ (w.l.o.g., we assume $j = 1$), exceeds that of the agents supporting any other opinion by a sufficiently-large *bias*, though the initial plurality itself might be very far from absolute majority. Our goal is to provide a dynamics that, with high probability, brings the system into the configuration in which all agents support the (initial) plurality opinion.

In this chapter we analyze the synchronous version of the dynamics introduced in [**AAE08**] and [**PVV09**], in the (uniform) $\mathcal{PULL}$ model: in every round, each agent pulls the opinion of a randomly-selected neighbor. If this opinion differs from its own, the agent enters the *undecided* state, an extra state that an agent can support. When an agent is in the undecided state and pulls an opinion, she gets that opinion. Finally, an agent that pulls

either the undecided opinion or its own opinion remains in its current state (see also Algorithm 2 and Table 1). Observe that, differently from other dynamics (e.g., the 3-Majority dynamics considered in Chapter 5), after the first round agents can also enter an undecided state, to which no opinion is associated.

We describe the notation that we adopt, part of which has already been introduced in the previous section. At each round t, the global state of the system is completely characterized by the corresponding opinion configuration, namely by the vector $\mathbf{c}^{(t)} = (c_1^{(t)}, c_2^{(t)}, \ldots, c_k^{(t)}, q^{(t)})$, where $c_i^{(t)}$ (respectively $q^{(t)}$) denotes the number of nodes having opinion i (respectively are in the undecided state) at the end of the t-th round. In the initial state, we always have $q^{(0)} = 0$. Given any initial opinion configuration $\mathbf{c} = (c_1, c_2, \ldots, c_k, 0)$, let us assume w.l.o.g. that $c_i \geqslant c_{i+1}$ for any $i \leqslant k-1$. Lower-case letters are used to denote functions of the observed opinion configuration at any specified time. Upper-case letters instead denote *random variables* (r.v.s). In particular, $Q^{(t)}$ and $C_i^{(t)}$ denote the number of nodes that are undecided and that have opinion i, respectively, at time t. At any time $t \geqslant 0$, the execution of the protocol (uniquely) determines the probability distribution of the (vectorial) random variable indicating the state at time t: $\mathbf{C}^{(t)} = (C_1^{(t)}, C_2^{(t)}, \ldots, C_k^{(t)}, Q^{(t)})$. Since we are considering complete graphs, this random process is clearly a finite-state Markov chain. To simplify notation, we omit the dependence of the random state on the initial opinion configuration. Since in the analysis presented in this chapter we don't need to condition on more complicated events than "$\mathbf{C}^{(t)} = \mathbf{c}''$, we simply write $\Pr(\cdot \mid \mathbf{c})$ in place of $\Pr(\cdot \mid \mathbf{C}^{(t)} = \mathbf{c})$, and similarly for expected values. Finally, when we condition the system to be in a fixed state $\mathbf{c}$ at some round, the random community sizes in the next round are denoted by C_i' and Q'.

6.1.0.1. *Global bias.* We define a *distance*[1] between opinion configurations as follows:

$$d\left(\mathbf{c}, \mathbf{c}'\right) = \sum_i \left(\frac{c_i}{c_1} - \frac{c_i'}{c'_1}\right)^2$$

In particular, consider the set M of the k possible *monochromatic* opinion configurations. For any $\mathbf{c}$, let

$$d(\mathbf{c}, M) = \min_{\mathbf{c}' \in M} \{d\left(\mathbf{c}, \mathbf{c}'\right)\}.$$

Definition 8 (Monochromatic Distance). Given an opinion configuration $\mathbf{c}$, its monochromatic distance is defined as

$$\mathsf{md}(\mathbf{c}) \;=\; \sum_{i=1}^{k} \left(\frac{c_i}{c_1}\right)^2,$$

[1]Note that $d\left(\bar{\mathbf{c}}, \bar{\mathbf{c}}'\right)$ is not a distance in the strict sense. See Section 6.4.1 for a formal discussion of this notion.

where c_1 is (one of) the plurality opinion(s).

It is easy to see that $\mathsf{md}(\mathbf{c}) = d(\mathbf{c}, M) + 1$.

6.2. High-level Analysis of the Undecided-State Dynamics

Generally speaking, when the initial configuration is sufficiently biased, the dynamics' evolution follows a typical pattern, characterized by well-distinct phases.

Understanding such a pattern requires a careful analysis. In this section, we provide an overview of this analysis, quantitatively describing a typical evolution of the process. We start from the expectations of a few key r.v.s

$$\mathbb{E}\left[C_i^{(t+1)} \,\middle|\, \mathbf{c}^{(t)}\right] = c_i^{(t)} \cdot \frac{c_i^{(t)} + 2q^{(t)}}{n} \tag{94}$$

$$\mathbb{E}\left[Q^{(t+1)} \,\middle|\, \mathbf{c}^{(t)}\right] = \frac{\left(q^{(t)}\right)^2 + \left(n - q^{(t)}\right)^2 - \sum_i \left(c_i^{(t)}\right)^2}{n} \tag{95}$$

These equations follow directly from the definition of the Undecided-State dynamics. From (94), we can appreciate the crucial role of the function $\frac{c_i^{(t)}+2q^{(t)}}{n}$: It represents the expected *growth rate* of every opinion community. The corresponding r.v. $C_i^{(t+1)} + 2Q^{(t+1)}$ is of particular interest when i is the plurality opinion[2]. In fact, a major novelty of our contribution is the discovery of a clean mathematical connection between the expected growth rate of the plurality and the monochromatic distance of the current configuration. The following expression formalizes this connection and plays a key role in our analysis. For every $t \geqslant 0$,

$$\mathbb{E}\left[C_1^{(t+1)} + 2Q^{(t+1)} \,\middle|\, \mathbf{c}^{(t)}\right] = \tag{96}$$

$$= \frac{n^2 + \left(n - 2q^{(t)} - c_1^{(t)})^2 + 2(R(\mathbf{c}^{(t)}\right) - \mathsf{md}(\mathbf{c}^{(t)}))\left(c_1^{(t)}\right)^2}{n},$$

where

$$R(\mathbf{c}) = \sum_{i=1}^{k} \frac{c_i}{c_1}.$$

Notice that $1 \leqslant \mathsf{md}(\mathbf{c}), R(\mathbf{c}) \leqslant k$ and $R(\mathbf{c}) \geqslant \mathsf{md}(\mathbf{c})$ (see (100)). The derivation of (96) becomes straightforward only after guessing the (non obvious) key role played by md as a measure of global bias. We observe that it is not linear in several parameters and its recursive form depends, through R and md, on the previous opinion configuration, as a whole. The resulting process evolution is thus rather complex and hard to analyze in a rigorous way (the details of this analysis can be found in Section 6.4). However, (96) allows us to informally characterize the main drivers of the process

[2]We are implicitly assuming that 1 remains the plurality opinion across the whole process. This holds w.h.p. under the assumptions of Theorem 9.

evolution. At the extremes, we have two complementary mechanisms that may determine an exponential (or quasi exponential) growth of C_1 and that qualitatively explain the leftmost (first phase) and rightmost (third phase) regions of Figure 20: Namely, large values of Q or of C_1 itself. In the latter case, growth follows a preferential attachment-like pattern. In the middle, we have a phase of relative "flat" growth that corresponds to Q dropping to a value close to $n/2$ and C_1 not being large enough to self-sustain an exponential growth. During this phase, growth is basically driven by the term $(R(\mathbf{c}) - \mathsf{md}(\mathbf{c}))c_1^2/n$, i.e., it crucially depends on the distance from the closest monochromatic configuration.

$$\mathbf{E}\left[c_i^{(t+1)} \,\middle|\, \mathbf{c}^{(t)}\right] = c_i^{(t)} \cdot$$

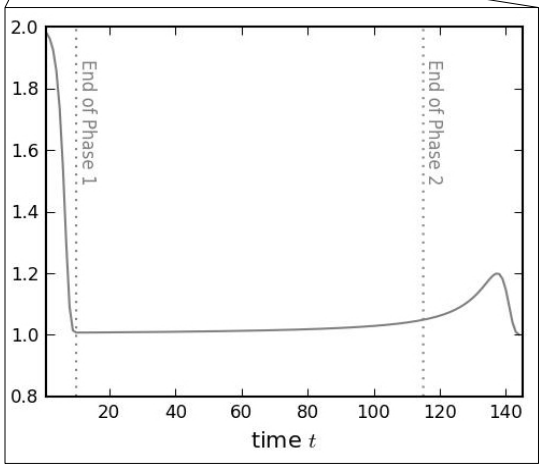

FIGURE 18. The different phases of the Undecided-State dynamics determined by the growth factor $\frac{C_1+2Q}{n}$ (cfr. Figure 20).

A further remark concerning (96) is that its proof crucially relies on properties of the plurality, the argument does not carry over to other opinions. In the next subsection, we give an overview of the analysis, deferring to Section 6.4 some major technical aspects which are mostly related to the rigorous characterization of phase-transition timings and the derivation of concentration bounds.

6.2.1. The process in a nutshell

The typical behaviour of the Undecided-State dynamics follows a characteristic pattern that exhibits three distinct phases, as exemplified in figures 20 and 18. Note that *the quantitative overview we provide below applies to typical evolutions*. We remark that the typical behavior holds w.h.p. under

the assumption that $\bar{c}_1 \geqslant (1+\alpha)\cdot \bar{c}_2$, where α is an arbitrarily small positive constant. Indeed this assumption guarantees that the initial plurality is preserved along the whole process, w.h.p.

6.2.1.1. *First round:* ***Rise of the undecided.*** The initial state is extremely unstable[3], since any node has a high probability of sampling a node of different opinion in the first round, ending up in the undecided state. Thus,

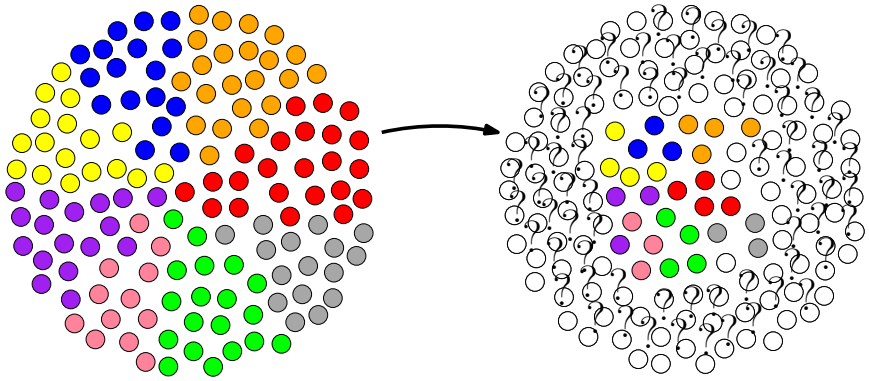

FIGURE 19. A representation of the first step of the Undecided-State Dynamics, where the size of each opinion i drops in expectation to $\frac{c_i^2}{n}$.

the first round sees dramatic changes in the system:

i) In general, a drastic drop in $C_i^{(1)}$'s (with "small"[4] ones simply disappearing w.h.p.);
ii) An explosive surge in $Q^{(1)}$, that possibly come to account for the vast majority;
iii) The initial plurality is preserved, w.h.p., though it drops in absolute terms.

A representation of this phase of the process is given in Figure 19.

Observe that, from (94) and (95) with $t=0$ and recalling that $q^{(0)}=0$, it follows that

$$\mathbb{E}\left[C_1^{(1)} \,\middle|\, \bar{\mathbf{c}}\right] = \frac{n}{R(\bar{\mathbf{c}})^2}\,,$$

$$\mathbb{E}\left[Q^{(1)} \,\middle|\, \bar{\mathbf{c}}\right] = n\left(1-\frac{1}{\Lambda(\bar{\mathbf{c}})}\right), \tag{97}$$

where

$$\Lambda(\bar{\mathbf{c}}) = \frac{R(\bar{\mathbf{c}})^2}{\mathsf{md}(\bar{\mathbf{c}})},$$

[3]Exceptions include cases that are less interesting, such as the one in which we have a strong absolute majority already at the onset.

[4]Namely, $o(\sqrt{n})$ in size.

and notice that $1 \leqslant \Lambda(\bar{\mathbf{c}}) \leqslant k$ (see (101)). Furthermore, $C_1^{(1)}$ and $Q^{(1)}$ are concentrated around their expectations (see Lemma 40 in Section 6.4).

6.2.1.2. *First phase:* ***Age of the undecided.*** The first phase starts right after round 1. In this phase, the C_i's grow (almost) exponentially fast while Q decreases. The duration of this phase depends on $\Lambda(\bar{\mathbf{c}})$ (and not just the magnitude of the initial bias). Those facts are discussed in the proof of Claim 1 that highlights key properties of the process marking the end of the first phase (for rigorous statements see Lemmas 42 and 43 in Section 6.4).

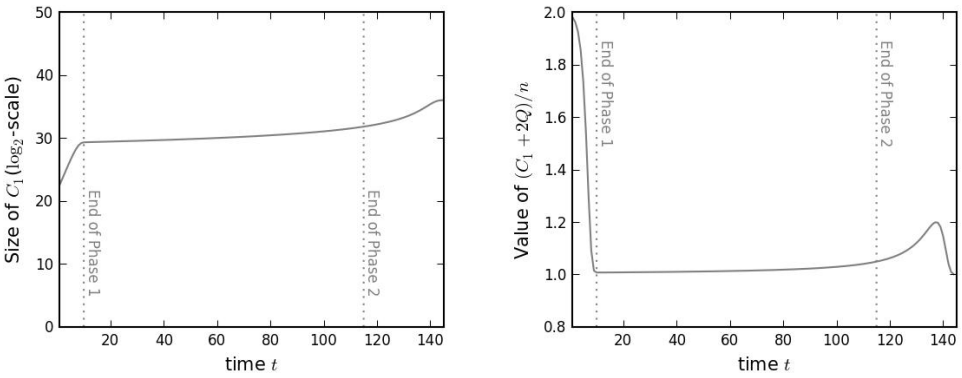

FIGURE 20. Typical evolution of the Undecided-State dynamics after the first round, for $n = 7 \cdot 10^{10}$ nodes and $k = (\frac{n}{\log n})^{\frac{1}{4}}$ opinions, with $c_1^{(0)} = 2\frac{n}{k}$ and $c_i^{(0)} = \frac{n}{k}\left(1 - \frac{2}{k}\right)$ for every $i \neq 1$.

CLAIM 1. *Within $T = O(\log \Lambda(\bar{\mathbf{c}}))$ rounds the system reaches a configuration such that w.h.p.*

$$Q^{(T)} = \frac{n}{2}\left(1 \pm \Theta\left(\frac{1}{\mathsf{md}(\bar{\mathbf{c}})}\right)\right),$$
$$C_1^{(T)} = \Theta\left(\frac{n}{\mathsf{md}(\bar{\mathbf{c}})}\right).$$

Furthermore, the relative ratios C_1/C_i are approximately preserved.

SKETCH OF PROOF. We first sketch the proof for the bound on Q. Assume that at some time t we are in a configuration $\mathbf{c}^{(t)}$ such that $q^{(t)} = (n/2)(1+\beta)$ for some $\beta > 0$. Notice that, choosing $\beta = 1 - \Theta(1/\Lambda(\bar{\mathbf{c}}))$, this assumption holds w.h.p. for $t = 1$ from the above overview of the first round. Then, from (95), we immediately have:

$$\mathbb{E}\left[Q^{(t+1)} \,\middle|\, \mathbf{c}^{(t)}\right] = \frac{n}{2}(1+\beta^2) - \frac{1}{n}\sum_i \left(c_i^{(t)}\right)^2.$$

Under reasonable assumptions on k, from the above inequality we have that w.h.p.

$$Q^{(t+1)} \leqslant (n/2)(1+\beta^2)$$

(see the proof of Lemma 43 in Section 6.4). Unfolding this argument for t rounds after round 1, we obtain that w.h.p.

$$Q^{(t+1)} \leqslant (n/2)(1+\beta^{2^t}).$$

Recalling that $\beta = 1 - \Theta(1/\Lambda(\bar{\mathbf{c}}))$, we thus obtain

$$Q^{(T)} \leqslant (n/2)\left(1+\Theta(1/\mathsf{md}(\bar{\mathbf{c}}))\right)$$

for

$$T = \log \Lambda(\bar{\mathbf{c}}) + O(\log\log \mathsf{md}(\bar{\mathbf{c}})).$$

Moreover, whenever

$$Q^{(t)} \geqslant (n/2)\left(1+\Theta(1/\mathsf{md}(\bar{\mathbf{c}}))\right),$$

we have w.h.p.

$$Q^{(t+1)} \geqslant (n/2)\left(1+\Theta(1/\mathsf{md}(\bar{\mathbf{c}}))\right),$$

which implies that w.h.p.

$$\left|Q^{(T)} - n/2\right| \leqslant \Theta\left(\frac{1}{\mathsf{md}(\bar{\mathbf{c}})}\right).$$

As for the claim for C_1, we next consider the evolution of the term $C_1^{(t)} + 2Q^{(t)}$ which, up to the factor $1/n$, determines the growth rate of $C_1^{(t+1)}$. Assume that $c_1^{(1)} + 2q^{(1)} = (1+\varepsilon)n$. We know from the analysis of the first round and in particular from (97), that this assumption holds w.h.p if we choose $\varepsilon \approx 1 - \Theta(1/\Lambda(\bar{\mathbf{c}}))$ (note that we are neglecting the contribution of $C_1^{(1)}$). Consequently, from (96) we get

$$\mathbb{E}\left[C_1^{(t+1)} + 2Q^{(t+1)} \,\middle|\, \mathbf{c}^{(t)}\right] \approx (1+\varepsilon^2)n.$$

Informally, by applying the argument above iteratively we obtain

$$\begin{aligned} C_1^{(2)} + 2Q^{(2)} &\approx (1+\varepsilon^2)n; \\ C_1^{(3)} + 2Q^{(3)} &\approx (1+\varepsilon^4)n; \\ &\cdots \\ C_1^{(t)} + 2Q^{(t)} &\approx (1+\varepsilon^{2^{t-1}})n. \end{aligned}$$

At this point, from (94) we get

$$\begin{aligned} C_1^{(t)} &\approx C_1^{(1)} \prod_{i=0}^{t-1} (1+\varepsilon^{2^i}) \approx C_1^{(1)} \prod_{i=0}^{t-1} \exp\left(\varepsilon^{2^i}\right) \\ &\approx C_1^{(1)} \exp\left(\sum_{i=0}^{t-1} \varepsilon^{2^i}\right) \\ &\approx C_1^{(1)} \exp\left(\sum_{i=0}^{t-1} \left(1 - \frac{1}{\Lambda(\bar{\mathbf{c}})}\right)^{2^i}\right). \end{aligned}$$

Since $T = \log \Lambda(\bar{\mathbf{c}}) + O(\log\log \mathsf{md}(\bar{\mathbf{c}}))$ it holds that

$$C_1^{(T)} \approx C_1^{(1)} \cdot \Theta\left(\Lambda(\bar{\mathbf{c}})\right) \approx \Theta\left(\frac{n}{\mathsf{md}(\bar{\mathbf{c}})}\right).$$

The last derivation follows from (97), which approximately holds w.h.p. (see also Lemma 40 in Section 6.4). gg

The proof outlined above highlights the following properties of the first phase:

i) The growth rate of plurality keeps "almost" exponential, while it quickly decreases mirroring the decrease of Q;
ii) The duration of the second phase is determined by $\log \Lambda(\bar{\mathbf{c}})$ (this can be as large as $\Theta(\log n)$ and as small as $O(1)$);
iii) From (97) it is possible to see that the factor $1/\mathsf{md}(\bar{\mathbf{c}})$, appearing in the expression of $C_1^{(T)}$ in the statement of Claim 1, corresponds to the fraction of the not-undecided nodes that belong to the plurality at the end of round 1.

6.2.1.3. *Second phase:* ***Plateau*** *or* ***Age of stability****.* The second phase is characterized by a slow increase of C_1, roughly at a rate $1+\Theta(1/\mathsf{md}(\bar{\mathbf{c}}))$ and a substantial stability of Q around the value $n/2$, as depicted in Figure 21. Indeed, if the system is in an opinion configuration $\mathbf{c}$ such that

$$q = \frac{n}{2}\left(1 \pm \Theta\left(\frac{1}{\mathsf{md}(\bar{\mathbf{c}})}\right)\right) \text{ and } c_1 = \Theta\left(\frac{n}{\mathsf{md}(\bar{\mathbf{c}})}\right).$$

(94) and (95) imply that

$$\begin{aligned} \mathbb{E}\left[Q' \,\middle|\, \mathbf{c}\right] &\approx \frac{n}{2}\left(1 - \Theta\left(\frac{1}{\mathsf{md}(\bar{\mathbf{c}})}\right)\right), \\ \mathbb{E}\left[C_1' \,\middle|\, \mathbf{c}\right] &\approx \left(1 + \Theta\left(\frac{1}{\mathsf{md}(\bar{\mathbf{c}})}\right)\right) c_1. \end{aligned}$$

By choosing the suitable constants we prove that the above relations hold w.h.p. (see Lemma 44 in Section 6.4). This is also the main argument for proving the following lower bound.

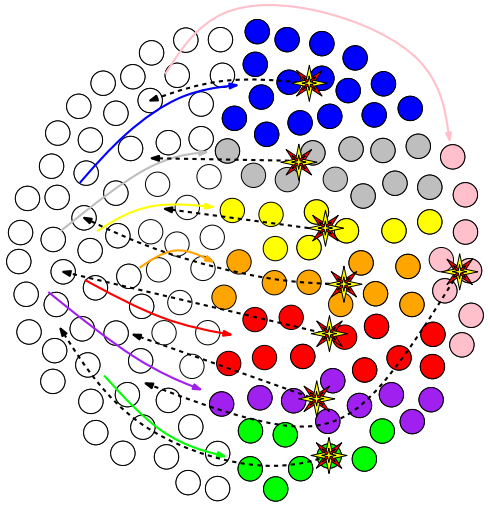

FIGURE 21. A representation of the plateau phase of the Undecided-State dynamics, where the number of agents which get a new opinion and that of new undecided ones almost balance each other. The "explosions" in the picture represent the event of a node seeing another one with a different opinion.

THEOREM 10 (Monochromatic Lower Bound). *Let* $k = O\left((n/\log n)^{1/6}\right)$. *Starting from any opinion configuration* $\mathbf{c}$ *the convergence time of the Undecided-State dynamics is* $\Omega(\mathsf{md}(\mathbf{c}))$, *w.h.p.*

However, as discussed above, since C_1 increases at a rate $1+\Theta(1/\mathsf{md}(\bar{\mathbf{c}}))$, after a plateau of $O(\mathsf{md}(\bar{\mathbf{c}})\log \mathsf{md}(\bar{\mathbf{c}}))$ rounds the system reaches a configuration $\mathbf{c}^{(t)}$ such that $R(\mathbf{c}^{(t)}) = 1 + o(1)$. This fact marks the end of the second phase, since the next phase yields a much faster growth of C_1. For a rigorous analysis of this part see Lemma 45 and Lemma 46 in Section 6.4.

6.2.1.4. *Third phase:* ***From plurality to totality***. Observe that, by definition of R, $C_1 = \frac{n-Q}{R}$ and, when the third phase starts, we have $R = 1+o(1)$: hence, $C_1 \approx n - Q$. Now, from (96), the leading term of the growth rate $\frac{C_1+2Q}{n}$ becomes $1 + (\frac{Q}{n})^2$. So, as long as Q is large (say $Q = \Theta(n)$), C_1 has an exponential growth while Q decreases. The above arguments, rigorously described in the proofs of Lemma 46 and Theorem 9 in Section 6.4, are the main ingredients to bound the time of the last phase. Finally, the whole analysis above yields the following upper bound.

THEOREM 9 (Monochromatic Upper Bound). *Let* $k = O\left((n/\log n)^{1/3}\right)$ *and let* $\mathbf{c}$ *be any initial configuration such that* $c_1 \geqslant (1+\alpha)\cdot c_2$ *where* α *is an arbitrarily small positive constant. Then within time* $O(\mathsf{md}(\mathbf{c})\cdot \log n)$ *the system converges to the plurality opinion, w.h.p.*

6.3. Extension on Expander Graphs

In this section we show how to adapt the Undecided-State dynamics to achieve plurality consensus on the class of d-regular expander graphs [**HLW06**] (with d denoting the degree of the nodes), at a polylogarithmic extra-cost in terms of local memory and time. The simple idea is to simulate the (uniform) random sampling of nodes' opinions by using n tokens, each originating at a different node and performing a (short) independent random-walk over the graph. It is well known [**LPW09**] that in every d-regular expander $G = (V, E)$ a lazy random walk has a uniform stationary distribution. Moreover, it is *rapidly mixing*, i.e., its mixing time is $\bar{t} = O(\log(1/\varepsilon) \log n)$ where ε is the desired bound on the total variation distance.

The modified dynamics works in synchronous *phases*, each of them consisting of exactly 2τ rounds (the suitable value for τ is defined later). During the first τ rounds a *forward process* takes place: Every node sends a token performing a random walk of at least $\bar{t}$-*hops* and thus sampling the opinion of a random node. A representation of this phase for a single node is given in Figure 22. In the next τ rounds we have a *backward process*: Every token is sent back to its source by "reversing" the path followed in the forward process.

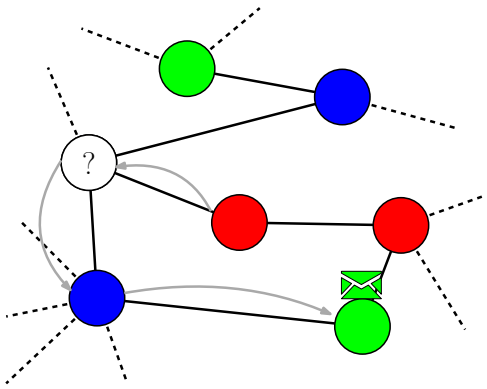

FIGURE 22. A representation of the first phase of the adaptation of the Undecided-State dynamics on expander graphs, in which each node sends a token performing a random walk of at least $\bar{t}$-*hops* and thus sampling the opinion of a random node.

If we were in the $\mathcal{LOCAL}$ model [**Pel00**], where each node can communicate with all its neighbors in one round, each phase of the above protocol would last exactly $2\bar{t}$ rounds. In the $\mathcal{GOSSIP}$ model [**CHHKM12**], each node can instead activate only one (bidirectional) link per round. Moreover, since we want *messages of limited size*, we assume that through each direction of an active link only one token can be transmitted.

We further assume that nodes enqueue tokens with a *FIFO* policy, breaking ties arbitrarily. The random walk performed by a token likely requires more than $\bar{t}$ rounds to perform (at least) $\bar{t}$ hops of the random walk, depending on the *congestion*, i.e. the maximum number of tokens in the queue of a node (see Figure 23 for a representation of the congestion issue). We thus need to bound the maximal congestion and use this bound, together with $\bar{t}$, to suitably set the right value for τ, so that every random walk is w.h.p. "mixed" enough. At time 2τ each node gets back its own token, and updates its state according to the Undecided-State dynamics. After that, a new phase starts, and the process iterates.

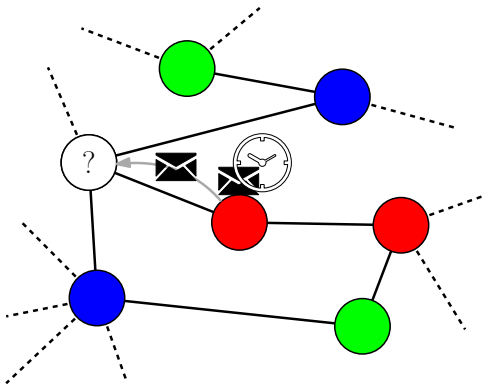

FIGURE 23. A representation of the congestion issue that arises if we try to perform many parallel random walks in the $\mathcal{GOSSIP}$ model with a FIFO policy: if two tokens are on the same node and have to move to different neighbors, one of them has to wait the next round to do that.

During the forward process, every token records the link labels of its random-walk and each node records, for any round, the (local) link label it has used (if any) to send a token at that round. Thanks to this information, every node can easily perform the backward process: At every round each node knows (if any) the neighbor it must contact to receive the right token back[5]. Notice that, since the backward process is perfectly specular to the forward one, the congestion is the same in both phases. Hence, both node memory and token message require $\Theta(\tau \log d)$ bits.

By setting a suitable value for τ, every token performs at least $\bar{t}$ hops, w.h.p. (some tokens may perform more hops than others). Thanks to the rapidly-mixing property, the opinion reported to the sender belongs to a random node, i.e., each node has probability $1/n \pm \varepsilon$ to be sampled (our analysis works setting $\varepsilon = O(1/n^2)$).

[5]Recall that in the $\mathcal{GOSSIP}$ model **[CHHKM12]**, agents can indeed contact one *arbitrary* neighbor per round.

In the next paragraph, we provide the main arguments of our congestion analysis (a formal analysis with all the details can be found in Section 6.4.7).

6.3.0.1. *Highlights on the congestion analysis.* Let $u \in [n]$ be a node, for every round $t \in [2\tau]$ of a phase, we consider the r.v. $\mathcal{Q}^{(t)}$ defined as the number of tokens in u at round t. Consider the number Y_t of tokens received by node u at round t (for brevity's sake, we omit index u in any r.v.). Then we can write $Y_t = \sum_{i\in[d]} X_{i,t}$ where $X_{i,t} = 1$ if the i-th neighbor of u sends a token to u and 0 otherwise. Observe that the r.v.s $X_{i,t}$ are not mutually independent. However, the crucial fact is that, for any t and any i, $\Pr(X_{i,t} = 1) \leqslant 1/d$, *regardless of the state of the system (in particular, independently of the value of the other r.v.s).* So, if we consider a family

$$\{\hat{X}_{i,t} \,:\, i \in [d],\ t \in [2\tau]\}$$

of i.i.d. Bernoulli r.v.s with $\Pr(\hat{X}_{i,t} = 1) = 1/d$, then Y_t is stochastically dominated by $\hat{Y}_t = \sum_{i\in[d]} \hat{X}_{i,t}$. For any node u and any round t, the r.v. $\mathcal{Q}^{(t)}$ is thus stochastically dominated by the r.v. $\hat{\mathcal{Q}}^{(t)}$ defined recursively as follows.

$$\begin{cases} \hat{\mathcal{Q}}^{(t)} &= \hat{\mathcal{Q}}^{(t-1)} + \hat{Y}_t - \chi_t \\ \hat{\mathcal{Q}}^{(0)} &= 1 \end{cases} \qquad \text{where} \quad \chi_t = \begin{cases} 1 & \text{if } \hat{\mathcal{Q}}^{(t-1)} > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Since our goal is to provide a concentration upper bound on $\mathcal{Q}^{(t)}$, we can do this by considering the "simpler" process $\hat{\mathcal{Q}}^{(t)}$. It turns out that "unrolling" $\hat{\mathcal{Q}}^{(t)}$ directly is far from trivial: we thus need the "right" way to write it by using only i.i.d. Bernoulli r.v.s. To this aim, for any $t \in [2\tau]$ and for any $s \in [t]$, we define the r.v.s

$$Z_{s,t} = \sum_{i=s}^{t} \hat{Y}_i - (t-s) \tag{98}$$

Informally speaking, $Z_{s,t}$ matches the value of $\hat{\mathcal{Q}}^{(t)}$ whenever $s \leqslant t$ was the last previous round s.t. $\hat{\mathcal{Q}}^{(s)} = 0$.

As a key fact (see the claim in the proof of Lemma 47 in the Section 6.4.7), we show that $\hat{\mathcal{Q}}^{(t)}$ can be written as a suitable function of $Z_{s,t}$ and χ_t so that it holds

$$\hat{\mathcal{Q}}^{(t)} \leqslant \max_{s\in[t]}\{Z_{s,t}\} \ \text{ and thus } \ \max_{t\in[2\tau]}\{\mathcal{Q}^{(t)}\} \leqslant \max_{s\leqslant t\leqslant 2\tau}\{Z_{s,t}\}. \tag{99}$$

From (98), the r.v. $Z_{s,t} + (t-s)$ is a sum of $d\cdot(t-s+1)$ i.i.d. Bernoulli r.v.s, each with expectation $1/d$. From the Chernoff bound (Lemma 76), it thus follows that, for constant $c > 0$ and any $1 \leqslant s \leqslant t \leqslant 2\tau$ we have

$$\Pr\left(Z_{s,t} \leqslant \max\left\{\sqrt{c(t-s+1)\log n},\, 3c\log n\right\}\right) \geqslant 1 - n^{-c/3}.$$

By taking the union bound over all $1 \leqslant s \leqslant t \leqslant 2\tau$, from the above bound and (99), we get the desired concentration bound on the maximal

node congestion during every phase:

$$\Pr\left(\max_{1\leqslant t\leqslant 2\tau}\mathcal{Q}^{(t)}\leqslant\max\left\{\sqrt{c\tau\log n},\,3c\log n\right\}\right)\geqslant 1-\frac{\tau^2}{n^{c/3}}.$$

The above congestion bound allows us to set the right value of τ, thus getting the following final result (its proof is given in Section 6.4.7).

THEOREM 11 (Monochromatic Bound on Expanders). *Let $G=(V,E)$ be a d-regular graph with constant expansion. For any initial configuration $\mathbf{c}$ such that the Undecided-State dynamics on the clique computes plurality consensus in $O(\mathsf{md}(\mathbf{c})\log n)$ rounds w.h.p., the modified Undecided-State dynamics computes plurality consensus on G in $O(\mathsf{md}(\mathbf{c})\,\mathrm{polylog}(n))$ rounds, w.h.p.*

REMARK 5. Notice that the analysis of the congestion also works in a scenario where every node generates a new token whenever its queue is empty, since it does not take care of the bound n on the overall number of nodes, and thus it is not tight.

6.4. Detailed Analysis of the Undecided-State Dynamics

In this section we work out the details of the analysis presented above. We begin with a closer look at the definition of the monochromatic distance.

6.4.1. The Monochromatic Distance

The results of this chapter highlight a fundamental dependence of convergence properties of the Undecided-State dynamics on a particular measure of the initial global bias. To mathematically characterize this we next introduce the following notion of distance between *equivalent* opinion configurations.

Given any opinion configuration $\mathbf{c}=(c_1,c_2,\ldots,c_k,q)$, consider the following ratio $R(\mathbf{c})=\sum_{i=1}^{k}c_i/c_1$. This allows us to define an equivalence relation $\equiv$ in the space $\mathcal{S}$

$$\mathbf{c}\equiv\mathbf{c}'\quad\text{iff}\quad R(\mathbf{c})=R(\mathbf{c}')$$

and the following function over pairs of equivalence classes (with an abuse of notation, for any opinion configuration $\mathbf{c}$, we denote its equivalence class as $\mathbf{c}$ as well)

$$d\left(\mathbf{c},\mathbf{c}'\right)\;=\;\sum_{i}\left(\frac{c_i}{c_1}-\frac{c'_i}{c'_1}\right)^2$$

The function $d(\cdot,\cdot)$ is a distance over the quotient space of $\mathcal{S}$. Let us now consider the equivalence class $\mathcal{M}$ of the k possible *monochromatic* opinion configurations and recall the definition of *monochromatic distance*.

DEFINITION 8 (Monochromatic Distance). Given an opinion configuration $\mathbf{c}$, its monochromatic distance is defined as

$$\mathsf{md}(\mathbf{c})\;=\;\sum_{i=1}^{k}\left(\frac{c_i}{c_1}\right)^2,$$

where c_1 is (one of) the plurality opinion(s).

It immediately follows that

$$\mathsf{md}\,(\mathbf{c}) \;=\; d(\mathbf{c},\mathcal{M})+1.$$

The simple considerations above entail that md defines a notion of distance from the monochromatic configuration that corresponds to the initial plurality. Consistently, it is straightforward to see that md is maximized by "uniform" configurations, i.e., configurations $\mathbf{c}$ such that $c_1 \approx n/k$. For every $\mathbf{c}$, it holds that

$$1 \;\leqslant\; R(\mathbf{c}), \mathsf{md}(\mathbf{c}) \;\leqslant\; k. \tag{100}$$

Finally, let us define the following ratio

$$\Lambda(\mathbf{c}) \;:=\; \frac{R(\mathbf{c})^2}{\mathsf{md}(\mathbf{c})}.$$

From the definitions of $R(\mathbf{c})$ and $\mathsf{md}(\mathbf{c})$ and from a simple application of the Cauchy-Schwartz inequality to $R(\mathbf{c})$, we get for every configuration $\mathbf{c}$

$$\Lambda(\mathbf{c}) \;\leqslant\; k. \tag{101}$$

6.4.2. General bounds on the Undecided-State dynamics

Before delving into the analysis, we provide some crucial properties that hold along the entire process. If $\mathbf{c} = (c_1, \ldots, c_k, q)$ is the current opinion configuration (i.e. state) of the Markov chain, then we can easily derive the following expected values of the next opinion configuration:

$$\mu_i = \mathbb{E}\left[C_i' \,\middle|\, \bar{\mathbf{c}}\right] = c_i \cdot \frac{c_i + 2q}{n} \quad (i \in [k]), \tag{102}$$

$$\begin{aligned}\mu_q = \mathbb{E}\left[Q' \,\middle|\, \bar{\mathbf{c}}\right] &= \frac{q^2 + \sum_{i\neq j} c_i \cdot c_j}{n} \\ &= \frac{q^2 + (n-q)^2 - \sum_i c_i^2}{n}.\end{aligned} \tag{103}$$

From (102), we can see the crucial role of the quantity $\frac{c_i+2q}{n}$: it represents the expected *growth rate* of every opinion community. The following lemma in fact formalizes such a connection by means of $R(\mathbf{c})$ and it plays a key role in the analysis of the entire process evolution. As we show in Lemma 41, $R(\mathbf{c})$ and $\mathsf{md}(\mathbf{c})$ are in fact strongly related.

Lemma 38 (Plurality Drift). *Assume that, at some round, the system is in an opinion configuration* $\mathbf{c}$ *such that* $c_1 \geqslant (1+\alpha)\, c_i$ *for any* $i \neq 1$ *and for some constant* $\alpha > 0$. *Then, at the next round, it holds that*

$$\mathbb{E}\left[\left.\frac{C_1' + 2Q'}{n}\,\right|\,\mathbf{c}\right] \;\geqslant\; 1 + \Gamma(\mathbf{c}),$$

where

$$\Gamma(\mathbf{c}) = \left(1 - \frac{c_1+2q}{n}\right)^2 + 2\,(1-\gamma)\,(R(\mathbf{c})-1)\left(\frac{c_1}{n}\right)^2,$$

with $\gamma = (1+\alpha)^{-1}$.

PROOF. Let $\beta = (1-\gamma)$. By using the hypothesis $c_1 \geqslant (1+\alpha)\, c_i$ we get

$$\mathsf{md}(\mathbf{c}) = \sum_i \frac{c_i^2}{c_1^2} \leqslant 1 + \frac{1}{(1+\alpha)} \sum_{i \neq 1} \frac{c_i}{c_1} = \gamma R(\mathbf{c}) + \beta.$$

Moreover, we can write q as $q = n - R(\mathbf{c})c_1$. Thanks to the above equations and (102) and (103), by simple manipulations, we get

$$\begin{aligned}
\mathbb{E}\left[\frac{C_1' + 2Q'}{n} \,\middle|\, \mathbf{c} \right] &= c_1 \cdot \frac{c_1 + 2q}{n^2} + 2\frac{q^2 + (n-q)^2 - \sum_i (c_i)^2}{n^2} \\
&= c_1 \cdot \frac{c_1 + 2q}{n^2} + 2\frac{q^2 + \left(R(\mathbf{c})^2 - \mathsf{md}(\mathbf{c})\right) \cdot (c_1)^2}{n^2} \\
&\geqslant c_1 \cdot \frac{c_1 + 2q}{n^2} + 2\frac{q^2 + \left(R(\mathbf{c})^2 - \gamma R(\mathbf{c}) - \beta\right) \cdot (c_1)^2}{n^2} \\
&= 1 + \left(1 - \frac{c_1 + 2q}{n}\right)^2 + 2\,(1-\gamma)\,(R(\mathbf{c}) - 1)\,\frac{c_1^2}{n^2}.
\end{aligned}$$

gg

Another useful property that is often used in our analysis is the fact that some crucial r.v.s are essentially monotone along the entire process. In the next lemma, we prove this monotonicity for the r.v.s $R(\mathbf{C}')$ and the ratios C_i'/C_1' (for $i \neq 1$).

LEMMA 39 (Monotonicity). *Assume that, at some round, the system is in an opinion configuration* $\mathbf{c}$ *such that, for some constant* $\alpha > 0$ *and a large enough constant* $\lambda > 0$ *it holds*

$$c_1 \geqslant (1+\alpha)\, c_i \;\; \textit{for any } i \neq 1 \quad \textit{and} \quad \mu_1 \geqslant \lambda \log n.$$

Then, at the next round, it holds w.h.p.

$$R(\mathbf{C}') < R(\mathbf{c}) \cdot \left(1 + O\left(\sqrt{\frac{\log n}{\mu_1}}\right)\right), \tag{104}$$

$$C_1' \geqslant (1+\alpha) \cdot C_i' \cdot \left(1 - O\left(\sqrt{\frac{\log n}{\mu_1}}\right)\right).$$

PROOF. As for Claim (104), since $R(\mathbf{C}') = \frac{\sum_i C_i'}{C_1'}$, it suffices to bound, respectively, C_1' and $\sum_i C_i'$. By applying the Chernoff bounds ((191) and

(192) in Lemma 76) and by using the hypothesis $\mu \geqslant \mu_1 \geqslant \lambda \log n$ we get

$$\Pr\left(C'_1 \leqslant \mu_1 \cdot \left(1 - \sqrt{\frac{2a \cdot \log n}{\mu_1}}\right) \mid \mathbf{c}\right) \leqslant \frac{1}{n^a}, \tag{105}$$

$$\Pr\left(C'_1 \geqslant \mu_1 \cdot \left(1 + \sqrt{\frac{3a \log n}{\mu_1}}\right) \mid \mathbf{c}\right) \leqslant \frac{1}{n^a},$$

$$\Pr\left(\sum_i C'_i \geqslant \mu \cdot \left(1 + \sqrt{\frac{3a \log n}{\mu}}\right) \mid \mathbf{c}\right) \leqslant \frac{1}{n^a}, \tag{106}$$

for any constant $a \in \left(0, \frac{\lambda}{3}\right)$.

Let A be the event in (105), let B be the event in (106) and let A^c and B^c be their complimentary events, respectively. From the union bound it follows that $\mathbf{P}\left(A^c \cap B^c\right) \geqslant 1 - \frac{2}{n^a}$. Moreover, since the following inequality holds

$$\begin{aligned}\frac{1 + \sqrt{\frac{3a \log n}{\mu}}}{1 - \sqrt{\frac{2a \log n}{\mu_1}}} &\leqslant \frac{1 + \sqrt{\frac{3a \log n}{\lambda \log n}}}{1 - \sqrt{\frac{2a \log n}{\lambda \log n}}} \\ &\leqslant 1 + \sqrt{\frac{ba \log n}{\lambda \log n}} \quad \text{with} \quad b = \left(\frac{\sqrt{3} - \sqrt{2}}{1 - \frac{3\sqrt{2a}}{\lambda}}\right)^2,\end{aligned}$$

we have that

$$\begin{aligned}&\Pr\left(R(\mathbf{C}') = \frac{\sum_i C'_i}{C'_1} < \frac{\sum_i c_i}{c_1} \cdot \left(1 + \sqrt{\frac{ba \log n}{\mu}}\right) \mid \mathbf{c}\right) \\ &\geqslant \Pr\left(\frac{\sum_i C'_i}{C'_1} < \frac{\sum_i c_i \cdot (c_i + q)}{c_1 \cdot (c_1 + q)} \cdot \left(1 + \sqrt{\frac{ba \log n}{\mu}}\right) \mid \mathbf{c}\right) \\ &= \Pr\left(\frac{\sum_i C'_i}{C'_1} < \frac{\mu}{\mu_1} \cdot \left(1 + \sqrt{\frac{ba \log n}{\mu}}\right) \mid \mathbf{c}\right) \\ &\geqslant \Pr\left(\frac{\sum_i C'_i}{C'_1} < \frac{\mu \cdot \left(1 + \sqrt{\frac{3a \log n}{\mu}}\right)}{\mu_1 \cdot \left(1 - \sqrt{\frac{2a \log n}{\mu_1}}\right)} \mid \mathbf{c}\right) \\ &\geqslant \mathbf{P}\left(A^c \cap B^c\right) \geqslant 1 - \frac{2}{n^a}.\end{aligned}$$

As for Claim (104), the hypothesis $c_1 \geqslant (1+\alpha)\, c_i$ clearly implies $\mu_1 \geqslant (1+\alpha)\cdot\mu_i$. Thus, by (105) we get

$$\begin{aligned}(107)\qquad &\Pr\left(C'_1 \leqslant (1+\alpha)\cdot\mu_i\cdot\left(1-\sqrt{\frac{2a\log n}{\mu_1}}\right)\;\middle|\,\mathbf{c}\right)\\ &\leqslant \Pr\left(C'_1 \leqslant \mu_1\cdot\left(1-\sqrt{\frac{2a\log n}{\mu_1}}\right)\;\middle|\,\mathbf{c}\right)\leqslant \frac{1}{n^a}.\end{aligned}$$

We now consider two cases. If $\mu_i < \mu_1/(6\,(1+\alpha))$ then, by the Chernoff bound ((189) in Lemma 76 with $\delta = \mu_1/(1+\alpha)$), with probability $1-n^{-\frac{\lambda}{1+\alpha}}$ it holds that $C'_i \leqslant \mu_1/((1+\alpha))$. Together with (105), this implies that w.h.p.

$$C'_1 > \mu_1\cdot\left(1-\sqrt{\frac{2a\log n}{\mu_1}}\right) > (1+\alpha)\, C'_i\cdot\left(1-\sqrt{\frac{2a\log n}{\mu_1}}\right).$$

On the other hand, if $\mu_i \geqslant \mu_1/(6\,(1+\alpha))$ then, from the Chernoff bound ((191) in Lemma 76) we get that

$$\begin{aligned}(108)\qquad &\Pr\left(C'_i \geqslant \mu_i\cdot\left(1+\sqrt{\frac{3a\log n}{\mu_i}}\right)\;\middle|\,\mathbf{c}\right)\\ &\leqslant \Pr\left(C'_i \geqslant \mu_i\cdot\left(1+\sqrt{\frac{3a\log n}{\mu_1/6(1+\alpha)}}\right)\;\middle|\,\mathbf{c}\right)\leqslant \frac{1}{n^a},\end{aligned}$$

for any $a \in \left(0, \frac{\lambda}{18(1+\alpha)}\right)$. Thus, by using (107), (108) and Fact 1 we get that w.h.p.

$$C'_1 \geqslant (1+\alpha)\cdot C'_i\cdot\left(1-O\left(\sqrt{\frac{\log n}{\mu_1}}\right)\right).$$

gg

6.4.3. First Round: *Rise of the undecided*

After the first round, a strong decrease of the opinion communities happens, while the undecided community gets to a large majority of the agents.

The next lemmas provide some formal statements about this behaviour which represent the key start-up of the process (and its analysis).

We implicitly assume that the process starts in a fixed initial opinion configuration $\mathbf{c} = (c_1, c_2, \ldots, c_k)$. So, in the next lemmas, events and related probabilities are conditioned on some fixed $\mathbf{c}$.

We observe that when k is large, i.e. when $k = \omega\left(n^b\right)$ for some $b \in (\frac{1}{2}, 1]$, if the process starts from "almost-uniform" opinion configurations then, after the first round, even the plurality may disappear, w.h.p.: indeed, if we consider any $\mathbf{c}$ such that $c_1 = O\left(\frac{n}{k}\right)$, then a simple application of the Markov inequality implies that $C'_1 = 0$, w.h.p. We thus focus on ranges of k such that $k < \sqrt{n/\log n}$.

LEMMA 40. *Let $k = o(\sqrt{n/\log n})$. Given any initial opinion configuration $\mathbf{c}$, after the first round it holds w.h.p.*

$$\begin{array}{rcccl} \frac{1}{2}\frac{n}{R(\bar{\mathbf{c}})^2} & \leqslant & C_1' & \leqslant & 2\frac{n}{R(\bar{\mathbf{c}})^2}, \\ n\left(1-\frac{2}{\Lambda(\bar{\mathbf{c}})}\right) & \leqslant & Q' & \leqslant & n\left(1-\frac{1}{2\Lambda(\bar{\mathbf{c}})}\right). \end{array}$$

PROOF. From (102) and recalling that in the initial configuration $q = 0$, we get

$$\mu_1 = \frac{(\bar{c}_1)^2}{n} = \frac{n}{R(\mathbf{c})^2}.$$

Similarly, from (103) we get

$$\mu_q = \frac{n^2 - \sum_i (\bar{c}_i)^2}{n} = \frac{n^2 - \mathsf{md}\cdot(\bar{c}_1)^2}{n} = n\left(1-\frac{1}{\Lambda(\bar{\mathbf{c}})}\right),$$

where the second equality follows from the definition of md, while the third one from the definition of $R(\bar{\mathbf{c}})$ and from simple manipulations. Since we assumed $k \leqslant o(\sqrt{n/\log n})$ then we have that

$$\mu_q = \frac{n}{R(\bar{\mathbf{c}})^2} \geqslant \frac{n}{k^2} = \omega(\log n).$$

The above inequality allows us to apply the Chernoff bound (Lemma 76) and prove the first claim (i.e. that on C_1').

Similarly, from (101), it holds

$$\frac{n}{\Lambda(\bar{\mathbf{c}})} \geqslant \frac{n}{k}.$$

This allows us to apply the additive version of the Chernoff bound (Lemma 76) and prove the second claim (i.e that on Q'). ∎

The next lemma relates $R(\mathbf{c})$ to $\mathsf{md}(\bar{\mathbf{c}})$ after the first round.

LEMMA 41. *Let $k = o(\sqrt{n/\log n})$. Given any initial opinion configuration $\bar{\mathbf{c}}$, after the first round it holds w.h.p.*

$$R(\mathbf{C}^{(1)}) \;\leqslant\; \mathsf{md}(\bar{\mathbf{c}})\cdot(1+o(1)).$$

PROOF. By definition of plurality opinion, it holds that $c_1 > n/k$. Therefore, by the hypothesis on k and (102), we get $\mu_1 = \omega(\log n)$ and then, by using the Chernoff bounds (Lemma 76), we can get concentration bounds on both the numerator and the denominator of $R(\mathbf{C}^{(1)})$ (as we did in the proof of Lemma 39). Formally, we have that w.h.p.

$$R(\mathbf{C}^{(1)}) = \frac{\sum_i C_i^{(1)}}{C_1^{(1)}} \leqslant \frac{\mu}{\mu_1}\cdot(1+o(1)).$$

Observe that, since in the initial opinion configuration $q = 0$, it holds

$$\frac{\mu}{\mu_1} = \frac{\sum_i (\bar{c}_i)^2}{(\bar{c}_1)^2}.$$

It follows that w.h.p.

$$R(\mathbf{C}^{(1)}) \leqslant \frac{\mu}{\mu_1} \cdot (1 + o(1)) = \frac{\sum_i (\bar{c}_i)^2}{(\bar{c}_1)^2} \cdot (1 + o(1)) \\ = \mathsf{md} \cdot (1 + o(1)),$$

concluding the proof. ∎

6.4.4. First phase: *Age of the undecided*

In this phase, the undecided community rapidly decreases to a value close to $n/2$ while the plurality reaches a size close to $n/(2\mathsf{md})$. When this happens, the ratios C_i/C_1 and $R(\mathbf{c})$ essentially keep their initial values and Q decreases to a value very close to $n/2$. The length of this phase is at most logarithmic.

The next lemma formalizes the aspects of this phase that are used to get the upper bound on the convergence time of the process.

LEMMA 42. *Let $k = o(\sqrt{n/\log^2 n})$ and let ε be any constant in $(0, \frac{1}{2})$. Let $\bar{\mathbf{c}}$ be any initial configuration such that, for any $j \neq 1$ and for some arbitrarily small constant $\alpha > 0$, $c_1 \geqslant (1+\alpha) \cdot c_j$. Then at some round $\tilde{t} = O(\log n)$ the process reaches a configuration $\mathbf{C}^{(\tilde{t})}$ such that w.h.p.*

$$\begin{cases} C_1^{(\tilde{t})} \geqslant \left(\frac{1}{16} - \frac{\varepsilon}{8}\right) \frac{n}{R(\mathbf{C}^{(\tilde{t})})}, & (109) \\ R(\mathbf{C}^{(\tilde{t})}) \leqslant \mathsf{md} \cdot (1 + o(1)), & (110) \\ C_1^{(\tilde{t})} \geqslant \left(1 + \frac{\alpha}{2}\right) \cdot C_i^{(\tilde{t})} \text{ for any opinion } i \neq 1, & (111) \\ \frac{C_1^{(\tilde{t})} + 2Q^{(\tilde{t})}}{n} > 1 + \frac{\varepsilon^2}{4}. & (112) \end{cases}$$

PROOF. We prove one claim at a time.

Proof of (109)**.** Let $\tilde{\varepsilon}$ be any positive constant in $(\varepsilon/2, \varepsilon)$. Two cases may arise. If $\bar{c}_1 > \left(\frac{1}{4} - \frac{\tilde{\varepsilon}}{2}\right) \cdot n$, by applying the Chernoff bound ((191) in Lemma 76) on the expected value of $C_1^{(1)}$ and using (100), it is easy to see that w.h.p.

$$C_1^{(1)} \geqslant \left(\frac{1}{16} - \frac{\varepsilon}{8}\right) n \geqslant \left(\frac{1}{16} - \frac{\varepsilon}{8}\right) \frac{n}{R(\mathbf{C}^{(1)})}.$$

If instead $\bar{c}_1 \leqslant \left(\frac{1}{4} - \frac{\tilde{\varepsilon}}{2}\right) \cdot n$. From Lemma 40 at round $t = 1$ we have w.h.p.

$$Q^{(1)} \geqslant n \left(1 - \frac{2}{\Lambda(\bar{\mathbf{c}})}\right) \geqslant n \left(1 - \frac{2c_1}{n}\right) \geqslant \frac{n}{2} + \tilde{\varepsilon} \cdot n,$$

where we used that $\Lambda(\bar{\mathbf{c}}) \geqslant R(\bar{\mathbf{c}}) = n/\bar{c}_1$.

In the generic configuration $\mathbf{c}$, as long as $q \geqslant \frac{n}{2} + \tilde{\varepsilon} \cdot n$, from (102) we have

$$\mu_1 \geqslant c_1 \cdot \left(\frac{1}{2} + \tilde{\varepsilon}\right),$$

thus, by applying the Chernoff bound ((191) in Lemma 76), we see that C_1 grows exponentially fast, w.h.p.

It follows that we can consider the first round such that $\tilde{t} = O(\log n)$ and $Q^{(\tilde{t})} < \frac{n}{2} + \tilde{\varepsilon} \cdot n$. This implies that

$$n - Q^{(\tilde{t})} \geqslant \frac{n}{2} - \tilde{\varepsilon} \cdot n,$$

hence

$$C_1^{(\tilde{t})} = \frac{n - Q^{(\tilde{t})}}{R(\mathbf{C}^{(\tilde{t})})} \geqslant \frac{\frac{n}{2} - \tilde{\varepsilon} \cdot n}{R(\mathbf{C}^{(\tilde{t})})}.$$

This proves (109).

Proof of (110)**.** Observe that, since $\bar{c}_1 \geqslant \frac{n}{k}$, then from (102) and the Chernoff bound ((191) in Lemma 76) it holds $C_1^{(1)} = \omega(\log^2 n)$, w.h.p. As we have already shown in the proof of Claim (109), after the first round C_1 grows exponentially until round $\tilde{t}$. It follows that we can repeatedly apply Lemma 39 and, together with Lemma 41, we get that w.h.p.

$$R(\mathbf{C}^{(\tilde{t})}) \leqslant \mathsf{md} \cdot \left(1 + o\left(\frac{1}{\log n}\right)\right)^{\log n} \leqslant \mathsf{md} \cdot (1 + o(1)).$$

This proves (110).

Proof of (111)**.** Similarly to the previous Claim proof, the repeated application of Lemma 39 until round $\tilde{t}$ and Fact 1 implies that w.h.p.

$$\begin{aligned} C_1^{(\tilde{t})} &\geqslant (1+\alpha) \cdot C_i^{(\tilde{t})} \cdot \left(1 - o\left(\frac{1}{\log n}\right)\right)^{\log n} \\ &= (1+\alpha) \cdot C_i^{(\tilde{t})} \cdot (1 - o(1)) \geqslant \left(1 + \frac{\alpha}{2}\right) \cdot C_i^{(\tilde{t})}. \end{aligned}$$

This proves (111).

Proof of (112)**.** Since, by the definition of $\tilde{t}$, it holds $q^{(\tilde{t}-1)} \geqslant \frac{n}{2} + \tilde{\varepsilon}$, then by Lemma 38 we get that

$$\mathbb{E}\left[C_1^{(\tilde{t})} + 2Q^{(\tilde{t})} \,\middle|\, \mathbf{c}^{(\tilde{t}-1)}\right] \geqslant (1 + \tilde{\varepsilon}^2) \cdot n.$$

Observe that $\mathbb{E}\left[C_1^{(\tilde{t})} + 2Q^{(\tilde{t})} \,\middle|\, \mathbf{c}^{(\tilde{t}-1)}\right]$ can be written as the expected value of the sum of the following independent r.v.s: given an opinion configuration $\mathbf{c}^{(\tilde{t}-1)}$, for each node i

$$X_i = \begin{cases} 1 & \text{if node } i \text{ has opinion 1 at the next round,} \\ 2 & \text{if node } i \text{ is undecided at the next round.} \end{cases}$$

Then (112) is an easy application of the Chernoff bound ((191) in Lemma 76). ∎

From the state conditions achieved after the first round (see Lemma 40), the next lemma shows that, within $O(\log n)$ rounds, the process reaches a configuration where Q gets very close to $n/2$ and C_1 is still relatively small, w.h.p. In the next section, we prove (see Theorem 10) that this fact forces

the process to "wait" for a time period $\Omega\left(\mathsf{md}(\bar{\mathbf{c}})\right)$ before the plurality (re-)starts to grow rapidly. This is the key ingredient of the lower bound in Theorem 10.

LEMMA 43. *Let $k \leqslant \varepsilon \cdot (n/\log n)^{1/6}$ be the initial number of opinions, where $\varepsilon > 0$ is a sufficiently small positive constant. Let $\bar{\mathbf{c}}$ be the initial opinion configuration and let $\mathbf{c}^{(1)}$ be the opinion configuration after the first round. If it holds that:*

$$\frac{1}{2}\,\frac{n}{R(\bar{\mathbf{c}})^2} \;\leqslant\; c_1^{(1)} \;\leqslant\; 2\,\frac{n}{R(\bar{\mathbf{c}})^2},$$

$$n\left(1-\frac{2}{\Lambda(\bar{\mathbf{c}})}\right) \;\leqslant\; q^{(1)} \;\leqslant\; n\left(1-\frac{1}{2\Lambda(\bar{\mathbf{c}})}\right),$$

within the next $O(\log n)$ rounds there is a round $\bar{t}$ such that w.h.p.

$$C_1^{(\bar{t})} \leqslant \gamma\frac{n}{\mathsf{md}(\bar{c})} \;\; \textit{and} \;\; \left|Q^{(\bar{t})}-\frac{n}{2}\right| \;\leqslant\; 2\frac{\gamma^2}{\mathsf{md}(\bar{\mathbf{c}})},$$

where $\gamma > 0$ is a sufficiently large constant.

PROOF. First, we prove that if at an arbitrary round t the number of undecided nodes is

$$q = (1+\delta)(n/2) \quad \text{with} \quad \frac{1}{\mathsf{md}(\bar{\mathbf{c}})} \leqslant \delta \leqslant 1-\frac{1}{2\Lambda(\bar{\mathbf{c}})},$$

then at the next round it holds that $Q' \leqslant \left(1+\delta^2\right)(n/2)$, w.h.p. Indeed, if we replace $q = (1+\delta)(n/2)$ in (103), we get that the expected value of Q' at the next round is

$$\begin{aligned}\mu_q &= \frac{1}{n}\Big(\left((1+\delta)\frac{n}{2}\right)^2 + \left((1+\delta)\frac{n}{2}\right)^2 - \sum_{j=1}^{k}\left(c_j\right)^2\Big)\\ &= \left(1+\delta^2\right)\frac{n}{2} - \frac{1}{n}\sum_{j=1}^{k}\left(c_j\right)^2\end{aligned}$$

Now observe that

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{k}(c_j)^2 &\geqslant \frac{1}{n}k\left(\frac{n-q}{k}\right)^2 = \frac{n}{4k}(1-\delta)^2\\ &\geqslant \frac{n}{4k}\cdot\left(\frac{1}{2\Lambda(\bar{\mathbf{c}})}\right)^2 \geqslant \frac{n}{16k^3},\end{aligned}$$

where in the last inequality we used (101), that is $\Lambda(\bar{\mathbf{c}}) \leqslant k$.

Therefore, since Q' is a sum of independent Bernoulli r.v., from the Chernoff bound ((190) in Lemma 76 with $\lambda = 1/16k^3$) it follows that

$$\Pr\left(Q' \geqslant \left(1+\delta^2\right)\frac{n}{2}\,\middle|\,\mathbf{c}\right) \leqslant \exp\left(-\frac{n}{128k^6}\right) \leqslant n^{-1/\left(128\varepsilon^6\right)}, \tag{113}$$

where in the last inequality we used the hypothesis on k.

Now we show that the number Q of undecided nodes, while decreasing quickly, cannot jump over the whole interval

$$\left[\frac{n}{2} - 2\gamma^2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})},\ \frac{n}{2} + 2\gamma^2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})}\right].$$

Observe that the function $f(q) = q^2 + (n-q)^2$ has a minimum for $q = n/2$, therefore for any

$$q \geqslant \frac{n}{2} + 2\gamma^2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})}$$

it holds that

$$f(q) \geqslant f\left(\frac{n}{2} + 2\gamma^2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})}\right).$$

Hence if at some round t we have that

$$q \geqslant \frac{n}{2}\left(1 + \frac{4\gamma^2}{\mathsf{md}(\bar{\mathbf{c}})}\right) \quad \text{and} \quad c_1 \leqslant \gamma n/\mathsf{md}(\bar{\mathbf{c}}),$$

in (103) we get

$$\begin{aligned}
\mu_q &\geqslant \frac{1}{n}\left(\left(\frac{n}{2} + 2\gamma^2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})}\right)^2 + \left(\frac{n}{2} + 2\gamma^2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})}\right)^2 - \sum_{j=1}^{k} c_j^2\right) \\
&= \frac{n}{2} + 4\gamma^4\frac{n}{\mathsf{md}(\bar{\mathbf{c}})^2} - \frac{1}{n}\sum_{j=1}^{k}(c_j)^2 \\
&\geqslant \frac{n}{2} - \frac{1}{n}\sum_{j=1}^{k}(c_j)^2 = \frac{n}{2} - \frac{(c_1)^2\,\mathsf{md}(\bar{\mathbf{c}})}{n} \geqslant \frac{n}{2} - \gamma^2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})},
\end{aligned}$$

where in the last inequality we used that $c_1 \leqslant \gamma n/\mathsf{md}(\bar{\mathbf{c}})$. Since Q' is a sum of n independent Bernoulli r.v., from the Chernoff bound (Lemma 76) it follows that

$$\begin{aligned}
\Pr\left(Q' \leqslant n/2 - 2\gamma^2 n/\mathsf{md}(\bar{\mathbf{c}}) \,\middle|\, \mathbf{c}\right) &\leqslant \exp\left(-2\gamma^2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})^2}\right) \\
&\leqslant \exp\left(-2\gamma^2\frac{n}{k^2}\right) \\
&\leqslant \exp\left(-\Omega\left(n^{2/3}\right)\right).
\end{aligned} \tag{114}$$

From (113), we get that w.h.p.

$$Q^{(t)} \leqslant \left(1 + \delta^{2^t}\right)\frac{n}{2}. \tag{115}$$

Hence, within

$$\log\left(\Lambda(\bar{\mathbf{c}})\right) + O(\log\log \mathsf{md}(\bar{\mathbf{c}}))$$

rounds, the number Q of undecided nodes is below $(n/2)(1 + 4\gamma^2/\mathsf{md}(\bar{\mathbf{c}}))$ w.h.p. Moreover, from (114) it follows that in one of such rounds we have that w.h.p.

$$\left|Q - \frac{n}{2}\right| \leqslant 2\gamma^2/\mathsf{md}(\bar{\mathbf{c}}).$$

It remains to show that, during this time, the plurality C_1 does not increase from less $2n/R(\bar{\mathbf{c}})^2$ to more than $\gamma n/\mathsf{md}(\bar{\mathbf{c}})$.

From (102) and (115) it follows that, as long as $c_1 \leqslant \gamma n/\mathsf{md}(\bar{\mathbf{c}})$, the increasing rate of C_1 at round t is at most

$$1 + \delta^{2^t} + \frac{\gamma}{\mathsf{md}(\bar{\mathbf{c}})},$$

w.h.p. For the first $\log(\Lambda(\bar{\mathbf{c}}))$ rounds, we can bound the above increasing rate with 2. Thus, after $\log(\Lambda(\bar{\mathbf{c}}))$ rounds we get that the plurality is $C_1 \leqslant 2n/\mathsf{md}(\bar{\mathbf{c}})$, w.h.p. As for the next $O(\log\log \mathsf{md}(\bar{\mathbf{c}}))$ rounds, we have that the plurality is at most

$$\begin{aligned}
&2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})} \cdot \prod_{t=l}^{L} \left(1 + \delta^{2^t} + \frac{\gamma}{\mathsf{md}(\bar{\mathbf{c}})}\right) \leqslant \\
&\leqslant 2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})} \cdot \exp\left(\sum_{t=l}^{L} \left(\delta^{2^t} + \frac{\gamma}{\mathsf{md}(\bar{\mathbf{c}})}\right)\right) \\
&\leqslant 2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})} \cdot \exp\left(O(1) + \frac{\log\log \mathsf{md}(\bar{\mathbf{c}})}{\mathsf{md}(\bar{\mathbf{c}})}\right) \leqslant \gamma\frac{n}{\mathsf{md}(\bar{\mathbf{c}})},
\end{aligned}$$

w.h.p., where in the last inequality we need to choose γ sufficiently large. ▣

REMARK 6. The two lemmas above refer to some rounds $\tilde{t}, \bar{t} = O(\log n)$ in which the process lies in a state satisfying certain properties. We observe that the analysis does never combine the two lemmas and thus it does not require that $\tilde{t} = \bar{t}$, indeed the first lemma is used to get the upper bound while the second one to get the lower bound on the convergence time. However, it is possible to prove that there is in fact a time interval (at the end of Phase 2) where both claims of the lemmas hold w.h.p.

6.4.5. Second phase: *Plateau* or *Age of stability*

This phase is characterized by a slow increase of c_1, roughly at a rate $1 + \Theta(1/\mathsf{md}(\bar{\mathbf{c}}))$. This fact is formalized in the next lemma and it is used to derive the lower bound on the convergence time of the process in Theorem 10.

LEMMA 44. *Let $\bar{\mathbf{c}}$ be the initial opinion configuration, let $k \leqslant \varepsilon\cdot(n/\log n)^{1/4}$ be the initial number of opinions, where $\varepsilon > 0$ is a sufficiently small positive constant. If there is a round $\bar{t}$ such that*

$$\left|q^{(\bar{t})} - \frac{n}{2}\right| \leqslant 2\gamma^2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})} \qquad \text{and} \qquad c_1^{(\bar{t})} \leqslant \gamma(n/\mathsf{md}(\bar{\mathbf{c}})),$$

where γ is an arbitrary positive constant, then the plurality C_1 remains smaller than $2\gamma(n/\mathsf{md}(\bar{\mathbf{c}}))$ for the next $\Omega(\mathsf{md}(\bar{\mathbf{c}}))$ rounds, w.h.p.

PROOF. Let us define $\delta = q - n/2$ and let Δ' be the random variable $Q' - n/2$ in the next round. From (102) we get

$$\mathbb{E}\left[\Delta' \,\middle|\, \mathbf{c}\right] = \frac{1}{n}\left(2\delta^2 - \sum_{j=1}^{k}(c_j)^2\right), \tag{116}$$

$$\mu_i = \left(1 + \frac{2\delta + c_i}{n}\right) c_i. \tag{117}$$

We show that, if

$$\delta \in \left(-\frac{2\gamma^2 n}{\mathsf{md}(\bar{\mathbf{c}})}, \frac{2\gamma^2 n}{\mathsf{md}(\bar{\mathbf{c}})}\right) \quad \text{and} \quad c_1 \leqslant \frac{2\gamma n}{\mathsf{md}(\bar{\mathbf{c}})},$$

then the increasing rate of C_1 is smaller than $(1+\Theta(1/\mathsf{md}(\bar{\mathbf{c}})))$, w.h.p. More precisely, we prove that w.h.p.

$$\begin{cases} |\delta| & \leqslant & 2\gamma^2 \frac{n}{\mathsf{md}(\bar{\mathbf{c}})} \\ c_1 & \leqslant & 2\gamma \frac{n}{\mathsf{md}(\bar{\mathbf{c}})} \end{cases} \Longrightarrow \begin{cases} |\Delta'| & \leqslant & 2\gamma^2 \frac{n}{\mathsf{md}(\bar{\mathbf{c}})} \\ C_1' & \leqslant & \left(1 + \frac{2\gamma(\gamma+1)+1}{\mathsf{md}(\bar{\mathbf{c}})}\right) c_1 \end{cases}.$$

As for the increasing rate of the plurality, from (117) it follows that

$$\begin{aligned} \mu_1 &= \left(1 + \frac{2\delta + c_1}{n}\right) c_1 \\ &\leqslant \left(1 + \frac{2\gamma^2 n/\mathsf{md}(\bar{\mathbf{c}}) + 2\gamma n/\mathsf{md}(\bar{\mathbf{c}})}{n}\right) c_1 \\ &= \left(1 + \frac{2\gamma(\gamma+1)}{\mathsf{md}(\bar{\mathbf{c}})}\right) c_1 \end{aligned}$$

Since C_1' can be written as a sum of $q + c_1 \leqslant n$ independent Bernoulli random variables, from the Chernoff bound ((190) in Lemma 76 with $\lambda = c_1/(n\mathsf{md}(\bar{\mathbf{c}}))$) it follows that

$$\begin{aligned} &\Pr\left(C_1 \geqslant \left(1 + \frac{2\gamma(1+\gamma)+1}{\mathsf{md}(\bar{\mathbf{c}})}\right) c_1 \,\middle|\, \mathbf{c}\right) \leqslant \\ &\leqslant \exp\left(-\frac{2\left(c_1/\mathsf{md}(\bar{\mathbf{c}})\right)^2}{n}\right) \overset{(a)}{\leqslant} \exp\left(-\frac{2n}{9k^4}\right) \overset{(b)}{\leqslant} n^{-2/(9\varepsilon^4)}, \end{aligned} \tag{118}$$

where in (a) we used that

$$c_1 \geqslant n - q/k \geqslant n/(3k) \quad \text{and} \quad \mathsf{md}(\bar{\mathbf{c}}) \leqslant k,$$

and in (b) we used the hypothesis $k \leqslant \varepsilon \cdot (n/\log n)^{1/4}$.

As for $\mathbb{E}\left[\Delta' \,|\, \mathbf{c}\right]$, according to (116), we have the upper bound

$$\mathbb{E}\left[\Delta' \,\middle|\, \mathbf{c}\right] \overset{(a)}{\leqslant} 2\frac{\delta^2}{n} \overset{(b)}{\leqslant} 8\gamma^4 \frac{n}{(\mathsf{md}(\bar{\mathbf{c}}))^2} \overset{(c)}{\leqslant} \gamma^2 \frac{n}{\mathsf{md}(\bar{\mathbf{c}})^2}, \tag{119}$$

where

- in (a) we discarded the non-negative term $\sum_{j=1}^{k}(c_j)^2$,

- in (b) we have used $|\delta| \leqslant 2\gamma^2 n/\mathsf{md}(\bar{\mathbf{c}})$, and
- in (c) we simply assumed that $\mathsf{md}(\bar{\mathbf{c}})$ is a sufficiently large constant, namely $\mathsf{md}(\bar{\mathbf{c}}) \geqslant 8\gamma^2$.

On the other hand, we have the lower bound

$$\mathbb{E}\left[\Delta' \,\middle|\, \mathbf{c}\right] = \frac{1}{n}\left(2\delta^2 - \sum_{j=1}^{k}(c_j)^2\right) \geqslant -\frac{1}{n}\sum_{j=1}^{k}(c_j)^2 \overset{(a)}{\geqslant} -\frac{k}{n}\left(\frac{n-q}{k}\right)^2 \overset{(b)}{\geqslant} -\frac{4}{9}\cdot\frac{n}{k} \overset{(c)}{\geqslant} -\frac{4}{9}\cdot\frac{n}{\mathsf{md}(\bar{\mathbf{c}})}, \tag{120}$$

where

- in (a) we used the fact that all c_j's are smaller than $n-q$,
- in (b) we used the fact that q is close to $n/2$, so $n-q$ is smaller than, say, $(2/3)n$, and finally
- in (c) we used the fact that $k \geqslant \mathsf{md}(\bar{\mathbf{c}})$.

Hence, from (119) and (120) we get

$$-\frac{4}{9}\frac{n}{\mathsf{md}(\bar{\mathbf{c}})} \leqslant \mathbb{E}\left[\Delta' \,\middle|\, \mathbf{c}\right] \leqslant \gamma^2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})}.$$

Since $\Delta' = Q' - n/2$ can be written as a sum of n independent random variables taking values $\pm 1/2$, from the appropriate version of Chernoff bound (Lemma 76) it thus follows that

$$\begin{aligned}\Pr\left(\Delta' \notin \left(-2\gamma^2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})},\, 2\gamma^2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})}\right) \,\middle|\, \mathbf{c}\right) \\ \leqslant \exp\left(-\Omega\left(\frac{n}{\mathsf{md}(\bar{\mathbf{c}})^2}\right)\right) \leqslant \exp\left(-\Omega\left(n^{1/2}\right)\right),\end{aligned} \tag{121}$$

where in the last inequality we used again the fact that $\mathsf{md}(\bar{\mathbf{c}}) \leqslant k \leqslant \varepsilon\,(n/\log n)^{1/4}$.

In order to formally complete the proof, let us now define event $\mathcal{E}_t = \mathcal{A}_t \wedge \mathcal{B}_t$, where $\mathcal{A}_t$ and $\mathcal{B}_t$ are the events

$$\mathcal{A}_t = \text{“}|\Delta^{(t)}| \leqslant 2\gamma^2\frac{n}{\mathsf{md}(\bar{\mathbf{c}})}\text{”},$$

$$\mathcal{B}_t = \text{“}C_1^{(t)} \leqslant \left(1 + \frac{2\gamma(1+\gamma)+1}{\mathsf{md}(\bar{\mathbf{c}})}\right)^t \cdot \gamma\frac{n}{\mathsf{md}(\bar{\mathbf{c}})}\text{”}.$$

Observe that

$$\left(1 + \frac{2\gamma(1+\gamma)+1}{\mathsf{md}(\bar{\mathbf{c}})}\right)^t \leqslant 2 \quad \text{for} \quad t \leqslant \frac{1}{4\gamma(1+\gamma)}\cdot\mathsf{md}(\bar{\mathbf{c}}).$$

Hence, if we set

$$T = \left\lfloor \frac{1}{4\gamma(1+\gamma)}\mathsf{md}(\bar{\mathbf{c}}) \right\rfloor,$$

from (118) and (121) it follows that, for every $j \in [\bar{t}, \bar{t}+T]$, we get

$$\Pr\left(\mathcal{E}_j \mid \cap_{i=1}^{j-1} \mathcal{E}_i\right) \geqslant (1-n^{-c}),$$

for a positive constant c that we can choose arbitrarily large. Thus, starting from the given opinion configuration $\mathbf{c}^{(\bar{t})}$, the probability that after T rounds the plurality $C_1^{(\bar{t}+T)}$ is at most $2\gamma n/\mathsf{md}(\bar{\mathbf{c}})$ is

$$\begin{aligned}
&\Pr\left(C_1^{(\bar{t}+T)} \leqslant 2\gamma\frac{n}{\mathsf{md}(\bar{\mathbf{c}})} \,\middle|\, \mathbf{c}^{(\bar{t})}\right)\\
&\geqslant \Pr\left(\bigcap_{j=\bar{t}}^{\bar{t}+T} \mathcal{E}_j\right) = \prod_{j=\bar{t}}^{\bar{t}+T} \Pr\left(\mathcal{E}_j \mid \bigcap_{i=\bar{t}}^{j-1} \mathcal{E}_i\right)\\
&\geqslant \left(1-n^{-c}\right)^T \geqslant 1 - Tn^{-c} \geqslant 1 - n^{-\Omega(1)}.
\end{aligned}$$

gg

THEOREM 10 (Monochromatic Lower Bound). *Let $k = O\left((n/\log n)^{1/6}\right)$. Starting from any opinion configuration $\mathbf{c}$ the convergence time of the Undecided-State dynamics is $\Omega(\mathsf{md}(\mathbf{c}))$, w.h.p.*

PROOF. From Lemma 40 and Lemma 43 it follows that there is a round $\bar{t}$, within the first $O(\log n)$ rounds, such that the process lies in an opinion configuration $\mathbf{c}^{(\bar{t})}$ where w.h.p.

$$\left|Q^{(\bar{t})} - n/2\right| \leqslant \frac{2\gamma^2}{\mathsf{md}(\bar{\mathbf{c}})} \quad \text{and} \quad C_1^{(\bar{t})} \leqslant \frac{\gamma n}{\mathsf{md}(\bar{\mathbf{c}})},$$

where γ is a sufficiently large constant. From Lemma 44, it then follows that the plurality C_1 remains smaller than $2\gamma(n/\mathsf{md}(\bar{\mathbf{c}}))$ for the next $\Omega(\mathsf{md}(\bar{\mathbf{c}}))$ rounds. gg

There is, however, a positive drift for the plurality working in this "long" phase as well: this minimal drift allows the process to reach a state which represents the end of this phase and from which the plurality can re-start to grow fast. In the next lemma we formally prove that the process exhibits the aforementioned minimal drift, while the latter phase-completion state is formalized in the subsequent Lemma 46.

LEMMA 45 (Minimal Drift). *Let $k = o\left(\sqrt{\frac{n}{\log n}}\right)$ and let $\varepsilon \in (0, \frac{1}{2})$ be an arbitrarily small positive constant. Given an opinion configuration $\mathbf{c}$ such that*

$$\begin{cases} c_1 \geqslant \beta \cdot \frac{n}{R(\mathbf{c})} & \text{for some constant } \beta > 0,\\ c_1 \geqslant (1+\alpha)\, c_i & \text{for some constant } \alpha > 0\\ & \text{and any } i \neq 1. \end{cases}$$

one of the following two holds, w.h.p.:

- *either*
$$R(\mathbf{C}') \leqslant 1 + \frac{\varepsilon}{3} \text{ and } Q' \leqslant \varepsilon n,$$
or

- *we have*

$$\frac{C_1' + 2Q'}{n} \geqslant 1 + \Omega\left(\frac{1}{R\left(\mathbf{c}\right)}\right).$$

PROOF. First, let us derive a lower bound on $C_1' + 2Q'$ that holds w.h.p. By Lemma 38

$$\mathbb{E}\left[C_1' + 2Q' \,\middle|\, \mathbf{c}\right] = n \cdot \left(1 + \Gamma(\mathbf{c})\right),$$

where

$$\Gamma(\mathbf{c}) = \left(1 - \frac{c_1 + 2q}{n}\right)^2 + 2\left(1-\gamma\right)\left(R(\mathbf{c}) - 1\right)\left(\frac{c_1}{n}\right)^2,$$

with $\gamma = (1+\alpha)^{-1}$. As in the proof of Lemma 42, observe that $\mathbb{E}\left[C_1' + 2Q' \mid \mathbf{c}\right]$ can be written as the expected value of the sum of the following independent r.v.s: given $\bar{\mathbf{c}}$, for each node i

$$X_i = \begin{cases} 1 & \text{if node } i \text{ has opinion 1 at round } t+1, \\ 2 & \text{if node } i \text{ is undecided at round } t+1. \end{cases}$$

Thus, we can apply the Chernoff bound ((191) in Lemma 76) to them and get that w.h.p.

$$C_1' + 2Q' \geqslant n \cdot \left(1 + \Gamma(\mathbf{c})\right)\left(1 - O\left(\sqrt{\frac{\log n}{n}}\right)\right). \tag{122}$$

Let us analyze (122) when $R(\mathbf{c}) > 1 + \frac{\varepsilon}{4}$ or $Q' > \frac{3}{4}\varepsilon n$.
If $R(\mathbf{c}) > 1 + \frac{\varepsilon}{4}$ we have that

$$\begin{aligned} \Gamma(\mathbf{c}) &\geqslant 2\left(1-\gamma\right)\left(R(\mathbf{c}) - 1\right)\left(\frac{c_1}{n}\right)^2 \\ &\geqslant 2\left(1-\gamma\right)\left(1 - \frac{1}{R(\mathbf{c})}\right)R(\mathbf{c}) \cdot \left(\frac{\beta}{R(\mathbf{c})}\right)^2 \\ &> \frac{\alpha\varepsilon\beta^2}{2(1+\alpha)(1+\varepsilon/4)} \cdot \frac{1}{R(\mathbf{c})}. \end{aligned} \tag{123}$$

On the other hand, if $R(\mathbf{c}) \leqslant 1 + \frac{\varepsilon}{4}$ then

$$c_1 = \frac{n-q}{R(\mathbf{c})} \geqslant \frac{n-q}{1+\varepsilon/4} \geqslant (n-q)(1-\varepsilon/4) \geqslant n - q - \frac{\varepsilon}{4}n,$$

hence, if it also holds that $q > \frac{3}{4}\varepsilon n$, the latter inequality implies that

$$1 - \frac{c_1 + 2q}{n} \leqslant \frac{\varepsilon}{4} - \frac{q}{n} \leqslant -\frac{\varepsilon}{2},$$

that is

$$\Gamma(\mathbf{c}) \geqslant \left(1 - \frac{c_1 + 2q}{n}\right)^2 \geqslant \frac{\varepsilon^2}{4}. \tag{124}$$

Therefore, if $R(\mathbf{c}) > 1 + \frac{\varepsilon}{4}$ or $q > \frac{3}{4}\varepsilon n$, then using (123), (124) and the given upper bound on the value of $R(\mathbf{c})$, from (122) we get

$$\begin{aligned}\frac{C_1' + 2Q'}{n} &\geqslant (1 + \Gamma(\mathbf{c}))\left(1 - O\left(\sqrt{\frac{\log n}{n}}\right)\right)\\ &\geqslant \left(1 + \frac{\sigma}{R(\mathbf{c})}\right)\left(1 - O\left(\sqrt{\frac{\log n}{n}}\right)\right) \geqslant \left(1 + \frac{\sigma}{2R(\mathbf{c})}\right),\end{aligned}$$

where

$$\sigma = \min\left\{\frac{\varepsilon^2}{4}R(\mathbf{c}), \frac{\alpha\varepsilon\beta^2}{2(1+\alpha)(1+\varepsilon/4)}\right\}.$$

It remains to show that if $R(\mathbf{c}) \leqslant 1 + \frac{\varepsilon}{4}$ and $q \leqslant \frac{3}{4}\varepsilon n$ then $R(\mathbf{C}') \leqslant 1 + \frac{\varepsilon}{3}$ and $Q' \leqslant \varepsilon n$, w.h.p.

In order to do so, observe that

$$\sum_{i \neq 1} c_i = (R(\mathbf{c}) - 1)c_1 \leqslant \frac{\varepsilon}{4}n.$$

It follows that

$$\begin{aligned}\mu_q &= \frac{q^2 + \sum_{i \neq j} c_i c_j}{n}\\ &\leqslant \frac{q^2 + 2c_1 \sum_{j \neq 1} c_j + \sum_{i \neq 1} c_i \sum_{j \neq 1} c_j}{n}\\ &\leqslant \left(\frac{3}{4}\varepsilon\right)^2 n + \frac{\varepsilon}{2}c_1 + \frac{\varepsilon^2}{16}n.\end{aligned}$$

Thanks to the Chernoff bound ((192) in Lemma 76) and since $\varepsilon < \frac{1}{2}$, the previous inequality implies that $Q' \leqslant \varepsilon n$, w.h.p. As for $R(\mathbf{C}')$, by applying Lemma 39 and using the Chernoff bound ((192) in Lemma 76), we get that $R(\mathbf{C}') \leqslant 1 + \frac{\varepsilon}{3}$, w.h.p. ∎

LEMMA 46. *Let $k = O((n/\log n)^{1/4})$ and let $\varepsilon > 0$ be an arbitrarily small constant. If the process is in an opinion configuration $\mathbf{c}^{(\tilde{t})}$ that satisfies the following conditions:*

$$\begin{cases}\dfrac{c_1^{(\tilde{t})} + 2q^{(\tilde{t})}}{n} = 1 + \Omega\left(\dfrac{1}{R(\mathbf{c}^{(\tilde{t})})}\right), & (125)\\ c_1^{(\tilde{t})} \geqslant \dfrac{1}{17}\dfrac{n}{R(\mathbf{c}^{(\tilde{t})})}, & (126)\\ R(\mathbf{c}^{(\tilde{t})}) = O(\mathsf{md}(\bar{\mathbf{c}})), & (127)\\ c_1^{(\tilde{t})} \geqslant (1+\alpha)\cdot c_i^{(\tilde{t})} \quad \textit{for some constant } \alpha > 0 & (128)\\ \qquad\qquad\qquad\qquad \textit{and for any opinion } i \neq 1,\end{cases}$$

then, after $T = O\left(\mathsf{md}(\bar{\mathbf{c}}) \cdot \log n\right)$ *rounds, the process is in an opinion configuration* $\mathbf{C}^{(\tilde{t}+T)}$ *such that w.h.p.*

$$\begin{cases} C_1^{(\tilde{t}+T)} \geqslant \dfrac{1}{17}\dfrac{n}{R(\mathbf{C}^{(\tilde{t}+T)})}, \\ R(\mathbf{C}^{(\tilde{t}+T)}) \leqslant 1 + \dfrac{\varepsilon}{3}, \\ Q^{(t+T)} \leqslant \varepsilon n, \\ C_1^{(\tilde{t}+T)} \geqslant (1+\alpha) \cdot C_i^{(\tilde{t}+T)}(1 - o(1)) \\ \qquad\qquad\qquad\qquad \textit{for any opinion } i \neq 1. \end{cases}$$

Proof. First, we show that, if we start in an opinion configuration $\mathbf{c}$ satisfying properties (125), (126), (127) and (128), then $\mathbf{C}'$ still satisfies the conditions (126), (127) and (128), w.h.p.

Using the Chernoff bound ((191) in Lemma 76) and conditions (126) and (125), we get that w.h.p.

$$\begin{aligned} C_1' &\geqslant \frac{c_1^{(\tilde{t})} + 2q^{(\tilde{t})}}{n} c_1 \left(1 - O\left(\sqrt{\frac{\log n}{\mu_1}}\right)\right) \\ &= \left(1 + \Omega\left(\frac{1}{R(\mathbf{c})}\right)\right) c_1 \geqslant \frac{1}{17}\frac{n}{R(\mathbf{c})}. \end{aligned}$$

In the first equality, we used that (125) and (126) together imply that w.h.p.

$$\mu_1 \geqslant c_1 \geqslant \frac{1}{17}\frac{n}{R(\mathbf{c})} \gg \frac{1}{R(\mathbf{c})},$$

which proves that $\mathbf{C}'$ also satisfies Condition (126), w.h.p. Moreover, Condition (126) allows us to apply Lemma 39 to get that w.h.p.

$$\begin{aligned} C_1' &\geqslant (1+\alpha) \cdot C_i' \cdot \left(1 - O\left((\log n/\mu_1)^{1/2}\right)\right), \\ R(\mathbf{C}') &< R(\mathbf{c}) \cdot \left(1 + O\left((\log n/\mu_1)^{1/2}\right)\right). \end{aligned}$$

proving that $\mathbf{C}'$ satisfies the hypotheses (127) and (128), w.h.p.

Now, by Lemma 45 and (127), it follows that either $R(\mathbf{C}') \leqslant 1 + \frac{\varepsilon}{3}$ and $Q' \leqslant \varepsilon n$, w.h.p. (in which case, we are done), or it holds w.h.p.

$$\frac{C_1' + 2Q'}{n} = 1 + \Omega\left(\frac{1}{R(\mathbf{c})}\right) = 1 + \Omega\left(\frac{1}{\mathsf{md}(\bar{\mathbf{c}})}\right).$$

In the latter case, $\mathbf{C}'$ satisfies also Condition (125) and the above argument can be iterated again. In particular, (125) implies that after $T = \Omega(\mathsf{md}(\bar{\mathbf{c}}) \log n)$ further rounds we have w.h.p.

$$\begin{aligned} C_1^{(\tilde{t}+T)} &= \left(1 + \Omega\left(\frac{1}{\mathsf{md}(\bar{\mathbf{c}})}\right)\right) c_1^{(\tilde{t}+T-1)} = \cdots = \\ &= \left(1 + \Omega\left(\frac{1}{\mathsf{md}(\bar{\mathbf{c}})}\right)\right)^T c_1^{(\tilde{t})} = n - o(n), \end{aligned}$$

and thus

$$R(\mathbf{C}^{(\tilde{t}+T)}) - 1 = \frac{\sum_{i\neq 1} C_i^{(\tilde{t}+T)}}{C_1^{(\tilde{t}+T)}} \leqslant \frac{\varepsilon}{3} \text{ and } Q^{(\tilde{t}+T)} \leqslant \varepsilon n.$$

□

6.4.6. Third phase: *From plurality to totality*

The next theorem connects the results achieved in the previous sections into a consistent picture, establishing an upper bound on the overall convergence time of the process. Its proof also highlights the main features of the final phase, during which plurality turns into the totality of agents at an exponential rate.

THEOREM 9 (Monochromatic Upper Bound). *Let $k = O\left((n/\log n)^{1/3}\right)$ and let $\mathbf{c}$ be any initial configuration such that $c_1 \geqslant (1+\alpha)\cdot c_2$ where α is an arbitrarily small positive constant. Then within time $O\left(\mathsf{md}(\mathbf{c})\cdot\log n\right)$ the system converges to the plurality opinion, w.h.p.*

PROOF. Let $\varepsilon > 0$ be an arbitrarily small positive constant. Thanks to Lemma 42, we can assume that at some time $\tilde{t} = O(\log n)$ the process reaches a configuration $\mathbf{C}^{(\tilde{t})}$ where it holds w.h.p.

$$\begin{cases} \dfrac{C_1^{(\tilde{t})} + 2Q^{(\tilde{t})}}{n} = 1 + \Omega\left(\dfrac{1}{R(\mathbf{c}^{(\tilde{t})})}\right), \\ C_1^{(\tilde{t})} \geqslant \dfrac{1}{17}\dfrac{n}{R(\mathbf{c}^{(\tilde{t})})}, \\ R(\mathbf{c}^{(\tilde{t})}) = O(\mathsf{md}), \\ C_1^{(\tilde{t})} \geqslant (1+\alpha)\cdot c_i^{(\tilde{t})}(1-o(1)) \quad \text{for any opinion } i \neq 1. \end{cases}$$

Assuming $\mathbf{c}^{(\tilde{t})}$, Lemma 46 determines the kick-off condition for a new phase in which both the undecided and the non-plurality opinion communities decrease exponentially fast. In particular, it implies that, within $O(\mathsf{md}\log n)$ further rounds, the process reaches a configuration $\mathbf{C}^{(t_{end})}$ such that it holds w.h.p.

$$\begin{cases} C_1^{(t_{end})} \geqslant \dfrac{1}{17}\dfrac{n}{R(\mathbf{c}^{(t_{end})})}, & (129) \\ C_1^{(t_{end})} \geqslant (1+\alpha)\cdot C_i^{(t_{end})}(1-o(1)) \quad \text{for any opinion } i \neq 1, & (130) \\ R(\mathbf{c}^{(t_{end})}) \leqslant 1 + \dfrac{\varepsilon}{3}, & (131) \\ Q^{t_{end}} \leqslant \varepsilon n. & (132) \end{cases}$$

Now, we show that starting from any configuration satisfying the conditions above, any community (including the undecided) other than the plurality decreases exponentially fast until disappearance. To this aim, let

$\psi = \sum_{i\neq 1} c_i + q$ and, as usual, let Ψ' be the r.v. associated to the value of ψ at the next time step. We prove that the following facts hold, w.h.p., in any round following t_{end}:

i) both Q and $\sum_{i\neq 1} C_i$ are bounded by quantities that decrease by a constant factor, so that at any time following t_{end}, Ψ is (upper) bounded by a quantity that decreases exponentially fast, thus $C_1 = n - \Psi$ is (lower) bounded by an increasing quantity;
ii) properties (130), still holds.

In the rest of this proof we assume $\varepsilon < 1/3$, which is consistent with the assumptions of Lemma 46.

To begin with, note that Property (131) implies $\sum_{i\neq 1} c_i \leqslant \frac{\varepsilon}{3} n$, so that

$$\sum_{i\neq j} c_i \cdot c_j \leqslant 2c_1 \sum_{j\neq 1} c_i + \sum_{i\neq 1} c_i \sum_{j\neq 1} c_j \leqslant \left(\frac{2}{3}\varepsilon + \frac{\varepsilon^2}{9}\right) n^2.$$

Therefore, properties (131) and (132) together imply

$$\begin{aligned} \mu_q &= \frac{(q)^2 + \sum_{i\neq j} c_i \cdot c_j}{n} \\ &\leqslant \left(\varepsilon^2 + \frac{2}{3}\varepsilon + \frac{\varepsilon^2}{9}\right) n < \frac{3}{4}\varepsilon n, \quad \text{and} \end{aligned} \tag{133}$$

$$\begin{aligned} \mathbb{E}\left[\sum_{i\neq 1} C_i' \,\middle|\, \mathbf{c}\right] &= \sum_{i\neq 1} \left(c_i \frac{c_i + 2q}{n}\right) \\ &\leqslant \frac{1}{3}\left(\frac{1}{3} + 2\right)\varepsilon^2 n = \frac{7}{9}\varepsilon^2 n < \frac{7}{27}\varepsilon. \end{aligned} \tag{134}$$

where we use the assumption that $\varepsilon < 1/3$. At this point, we can use the Chernoff bound ((192) in Lemma 76) to show that (133) and (134) hold, w.h.p. (up to a multiplicative factor $1 + o(1)$). This proves that both Q and $\sum_{i\neq 1} C_i$ (and hence Ψ) decrease by a constant factor in a round[6], w.h.p.

It remains to observe that, when q and/or $\sum_{i\neq 1} c_i$ become $O(\log n)$, an application of the Chernoff bound ((189) in Lemma 76) shows that they remain below this value in the subsequent rounds, w.h.p. This completes the proof of i).

Moreover, since $C_1' = n - \Psi'$, i) implies that C_1' is lower bounded by an increasing quantity, w.h.p. Additionally, property (129) and the just-proved i), together with property (130), imply the assumptions of Lemma 39, allowing us to show that property (130) still holds at the end of next round, w.h.p. As a consequence, we have that in at most $\tau = O(\log n)$

[6]In fact, a more careful analysis, unnecessary to prove the current result, could use (134) to show that $\sum_{i\neq 1} C_i$ decreases superexponentially fast.

rounds we reach an opinion configuration $\bar{C}^{(t_{end}+\tau)}$ such that w.h.p.

$$Q^{(t_{end}+\tau)} + \sum_{i \neq 1} C_i^{(t_{end}+\tau)} = O(\log n).$$

Finally, we can apply Markov's inequality on the value of $\sum_{i \neq 1} C_i^{(t_{end}+\tau)}$ to show that at the next round all opinion communities except for the plurality one disappear, w.h.p. ∎

6.4.7. Node congestion analysis

The parallel random walks described in Section 6.3 yield variable token queues in the nodes. Recall that, for each node $u \in [n]$, and for every round $t \in [2\tau]$ of the phase, we consider the r.v. $\mathcal{Q}_u^{(t)}$ defined as the number of tokens in u at round t of any phase of the modified dynamics. In the next lemma we prove a useful bound on the maximal congestion in a phase of length 2τ.

LEMMA 47. *Consider a phase of length $2\tau \geqslant 1$ of the above protocol on a d-regular graph $G = (V, E)$. Let $u \in V$ be any node and let t be any round of the phase. Then, for any constant $c > 0$, it holds that*

$$\Pr\left(\max_{1 \leqslant t \leqslant 2\tau} \mathcal{Q}_u^{(t)} \leqslant \max\left\{\sqrt{2c\tau \log n},\, 3c \log n\right\}\right) \geqslant 1 - \frac{(2\tau)^2}{n^{c/3}}.$$

PROOF. Consider the number Y_t of tokens received by a fixed node u at round t (for brevity's sake, we omit index u in any r.v.). Then we can write

$$Y_t = \sum_{i \in [d]} X_{i,t},$$

where $X_{i,t} = 1$ if the i-th neighbor of u sends a token to u and 0 o.w.. Observe (again) that the r.v.s $X_{i,t}$ are not mutually independent. However, the crucial fact is that, for any t and any i, it holds

$$\Pr(X_{i,t} = 1) \leqslant \frac{1}{d},$$

regardless the state of the system (in particular, independently of the value of the other r.v.s).

So, if we consider a family

$$\{\hat{X}_{i,t} \,:\, i \in [d],\ t \in [2\tau]\}$$

of i.i.d. Bernoulli r.v.s with $\Pr(\hat{X}_{i,t} = 1) = 1/d$, then Y_t is stochastically smaller than

$$\hat{Y}_t = \sum_{i=1}^{d} \hat{X}_{i,t}.$$

For any node u and any round t, the r.v. $\mathcal{Q}^{(t)}$ is thus stochastically smaller than the r.v. $\hat{\mathcal{Q}}^{(t)}$ defined recursively as follows.

$$\begin{cases} \hat{\mathcal{Q}}^{(t)} &= \hat{\mathcal{Q}}^{(t-1)} + \hat{Y}_t - \chi_t \\ \hat{\mathcal{Q}}^{(0)} &= 1 \end{cases} \qquad \text{where } \chi_t = \begin{cases} 1 & \text{if } \hat{\mathcal{Q}}^{(t-1)} > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Since our goal is to provide a concentration upper bound on $\mathcal{Q}^{(t)}$, we can do it by considering the "simpler" process $\hat{\mathcal{Q}}^{(t)}$. By the way, unrolling $\hat{\mathcal{Q}}^{(t)}$ directly is far from trivial: We need the "right" way to write it by using only i.i.d. Bernoulli r.v.s. Let us see how.

For any $t \in [2\tau]$ and for any $s \in [t]$, define the r.v.

$$Z_{s,t} = \sum_{i=s}^{t} \hat{Y}_i - (t-s). \tag{135}$$

Informally speaking, $Z_{s,t}$ matches the value of $\hat{\mathcal{Q}}^{(t)}$ whenever $s \leqslant t$ was the last previous round s.t. $\hat{\mathcal{Q}}^{(s)} = 0$. As a key-fact we show that $\hat{\mathcal{Q}}^{(t)}$ can be bounded by the maximum of $Z_{s,t}$ for $s \leqslant t$.

Claim 2. *For any $t \in [2\tau]$ it holds that*

$$\hat{\mathcal{Q}}^{(t)} \leqslant \max\{Z_{s,t} \,:\, s = 1, \ldots, t\},$$

and thus

$$\max\{\mathcal{Q}^{(t)} \,:\, 1 \leqslant t \leqslant 2\tau\} \leqslant \max\{Z_{s,t} \,:\, 1 \leqslant s \leqslant t \leqslant 2\tau\}. \tag{136}$$

Proof of the Claim. For any $s \in [t]$, let

$$\chi_{s,t} = \prod_{r=s}^{t} \chi_r$$

be the r.v. taking value 1 if $\hat{\mathcal{Q}}^{(r-1)} > 0$ for all $s \leqslant r \leqslant t$ and 0 otherwise. It is easy to prove by induction that $\hat{\mathcal{Q}}^{(t)}$ can be written as

$$\hat{\mathcal{Q}}^{(t)} = \sum_{s=2}^{t} (1-\chi_{s-1})\chi_{s,t} Z_{s-1,t} + \chi_{1,t} Z_{1,t} + (1-\chi_t) Z_{t,t}. \tag{137}$$

Since

$$\sum_{s=2}^{t} (1-\chi_{s-1})\chi_{s,t} + \chi_{1,t} = 1,$$

the sum in (137) is not larger than the maximum of the $Z_{s,t}$, hence

$$\hat{\mathcal{Q}}^{(t)} \leqslant \max\{Z_{s,t} \,:\, s = 1, \ldots, t\}$$

and

$$\max\{\mathcal{Q}^{(t)} \,:\, 1 \leqslant t \leqslant 2\tau\} \leqslant \max\{Z_{s,t} \,:\, 1 \leqslant s \leqslant t \leqslant 2\tau\}.$$

gg (of Claim 2)

Let us consider (135): The r.v. $Z_{s,t}+(t-s)$ is a sum of $d\cdot(t-s+1)$ i.i.d. Bernoulli r.v.s each one with expectation $1/d$. From the Chernoff bounds ((192) and (189) in Lemma 76), for any $1 \leqslant s \leqslant t$, it holds that

$$\Pr\left(Z_{s,t} \leqslant \max\left\{\sqrt{c(t-s+1)\log n},\, 6c\log n\right\}\right) \geqslant 1 - n^{-c/3}.$$

By taking the union bound over all $1 \leqslant s \leqslant t \leqslant 2\tau$, from the above bound and (136) we can get the desired concentration bound on the maximal node congestion during every phase:

$$\Pr\left(\max_{1\leqslant t\leqslant 2\tau} \mathcal{Q}^{(t)} \leqslant \max\left\{\sqrt{2c\tau\log n},\, 6c\log n\right\}\right) \geqslant 1 - \frac{(2\tau)^2}{n^{c/3}}.$$

gg

Let $\mathrm{t}^G_{\mathrm{mix}}(\varepsilon)$ be the first round such that the total variation distance between the simple random walk starting at an arbitrary node and the uniform distribution is smaller than ε, i.e.

$$\mathrm{t}^G_{\mathrm{mix}}(\varepsilon) = \inf\{t \in \mathbb{N} : \|P^t(u,\cdot) - \pi\| \leqslant \varepsilon \ \text{ for all } u \in V\}.$$

Notice that for any $\varepsilon > 0$ it holds that (see e.g. (4.36) in [**LPW09**])

$$\mathrm{t}^G_{\mathrm{mix}}(\varepsilon) \leqslant \log\left(\frac{1}{\varepsilon}\right) \mathrm{t}^G_{\mathrm{mix}}\left(\frac{1}{2e}\right). \tag{138}$$

As a consequence of the above Lemma, we can now set the right value of τ, thus getting the following result.

THEOREM 23 (Uniform $\mathcal{GOSSIP}$ Simulation on Expanders). *Let $G = ([n], E)$ be a d-regular graph with $t^G_{mix}(1/4) = \mathrm{polylog}(n)$. Each round of a protocol on the clique in the uniform $\mathcal{GOSSIP}$ model can be simulated on G in the $\mathcal{GOSSIP}$ model in* $\mathrm{polylog}(n)$ *rounds by exchanging messages of* $\mathrm{polylog}(n)$ *size, w.h.p.*

PROOF. Let $2\tau = \alpha\bar{t}^2\log n$ be the length of the phase, where $\bar{t} = \mathrm{t}^G_{\mathrm{mix}}\left(1/n^2\right)$ and α is a suitable constant that we fix later. From Lemma 47, we have that the maximum number of tokens in every node at any round of the phase is at most

$$\sqrt{2c\tau\log n} = \sqrt{\alpha c}\cdot\bar{t}\log n,$$

w.h.p Since tokens are enqueued with a FIFO policy, each single hop of the random walk performed by a token can be delayed for at most the above number of rounds. Hence, in order to perform $\bar{t}$ hops of the random walk, a token takes at most $\sqrt{\alpha c}\cdot\bar{t}^2\log n$ rounds, w.h.p.

By choosing $\alpha \geqslant 4c$ we have that this number is smaller than τ. This allows us to set τ so that the forward process and the backward one can both complete safely.

By union bounding over all tokens we thus have that during the phase all tokens perform at least $\bar{t}$ hops of a random walk and report back to the sender the opinion of the node they reached after $\bar{t}$ hops, w.h.p.

Finally, notice that from (138) it follows that $\bar{t} = \text{polylog}(n)$. The phase length and the size of the exchanged messages are thus $\text{polylog}(n)$ as well. ⊠

Since a lazy random walk on regular expanders (see e.g. [**HLW06**]) has $\text{polylog}(n)$ mixing time, from the above theorem and our result on the Undecided-State dynamics on the clique we get the following final result.

THEOREM 11 (Monochromatic Bound on Expanders). *Let $G = (V, E)$ be a d-regular graph with constant expansion. For any initial configuration $\mathbf{c}$ such that the Undecided-State dynamics on the clique computes plurality consensus in $O(\mathsf{md}(\mathbf{c}) \log n)$ rounds w.h.p., the modified Undecided-State dynamics computes plurality consensus on G in $O(\mathsf{md}(\mathbf{c}) \,\text{polylog}(n))$ rounds, w.h.p.*

CHAPTER 7

Congested Random Walks

In this chapter we study the parallel random walks process in the uniform $\mathcal{PUSH}$ model on a complete topology, proving the results presented in Section 2.4. We conveniently reformulate the previous process as the following *repeated balls-into-bins* process: n balls are initially assigned to n bins in an arbitrary way; In every subsequent round, from each non-empty bin one ball is chosen according to some fixed strategy (random, FIFO, etc), and re-assigned to one of the n bins uniformly at random (see Figure 24).

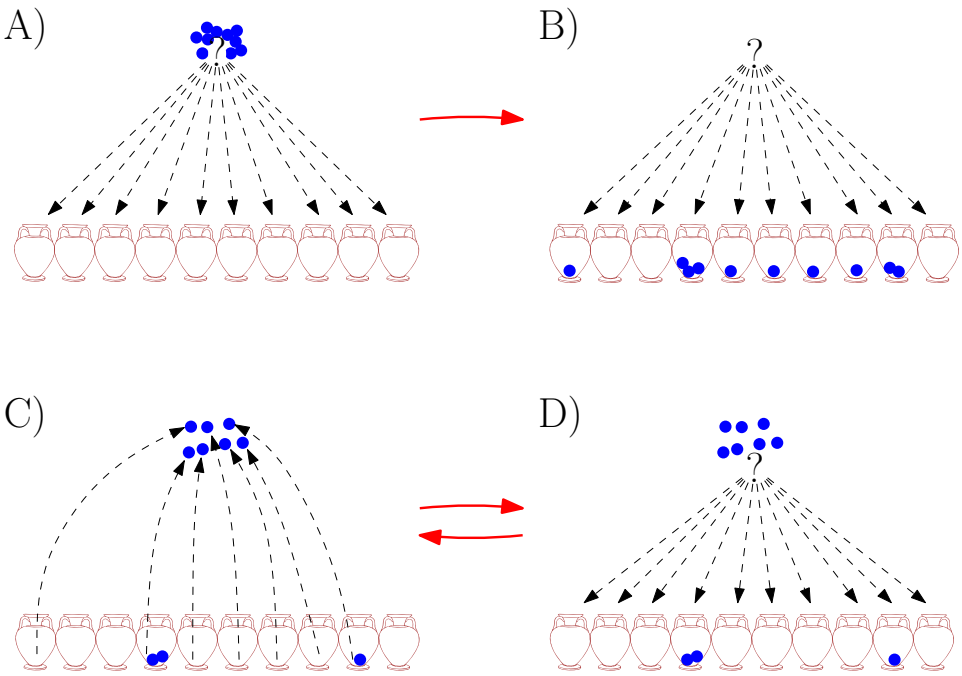

FIGURE 24. **A) and B):** In the balls-into-bins process, each ball is thrown in one bin chosen independently and u.a.r. **C) and D):** In the repeated balls-into-bins process, at each round we pick one ball from each non-empty bin (e.g. let us assume that the current configuration is the one in B)), and throw them again u.a.r.

Adopting the framework of (probabilistic) self-stabilization (Definition 9), we define a configuration *legitimate* if its maximum load is $\mathcal{O}(\log n)$. We prove that, starting from any configuration, the process converges to a legitimate configuration in linear time and then it takes on only legitimate configurations over a period of length bounded by *any* polynomial in n, w.h.p. This implies that the process is self-stabilizing and that every ball traverses all bins in $\mathcal{O}(n \log^2 n)$ rounds, w.h.p.

7.1. Self-Stabilization of repeated balls into bins

In order to study the maximum load of the repeated balls-into-bins process, the state of the system is completely characterized by the load of every bin. As in Section 6.3 (Chapter 6), for each bin $u \in [n]$ let $\mathcal{Q}_u^{(t)}$ be the r.v.s[1] indicating the number of balls, i.e. the *load*, in u at round t. We write $\mathbf{Q}^{(t)}$ for the vector of these random variables, i.e.

$$\mathbf{Q}^{(t)} = (\mathcal{Q}_u^{(t)} \ : \ u \in [n]).$$

We write $\mathbf{q} = (q_1, \ldots, q_n)$ for a *(load) configuration*, i.e., $q_u \in \{0, 1, \ldots, n\}$ for every $u \in [n]$ and $\sum_{u=1}^{n} q_u = n$. We define the *maximum load* of a configuration $\mathbf{q} = (q_1, \ldots, q_n)$ as

$$M(\mathbf{q}) \ = \ \max\{\, q_u \ : \ u \in [n] \,\},$$

and, for brevity' sake, given any round t of the process, we define

$$M^{(t)} \ = \ M(\mathbf{Q}^{(t)}).$$

According to the above definition, we say that a configuration $\mathbf{q}$ is *legitimate* if $M(\mathbf{q}) \leqslant \beta \cdot \log n$, for some absolute constant $\beta > 0$.

In this section we prove the main result of this chapter, which we prove in Section 7.1.4.

THEOREM 12 (Repeated Balls into Bins Max Load). *Let c be an arbitrarily-large constant and let $\mathbf{q}$ be any legitimate configuration. Let the* repeated balls-into-bins process *start from $\mathbf{Q}^{(0)} = \mathbf{q}$. Then, over any period of length $\mathcal{O}(n^c)$, the process visits only legitimate configurations, w.h.p., i.e. $M^{(t)} = \mathcal{O}(\log n)$ for all $t = \mathcal{O}(n^c)$, w.h.p. Moreover, starting from any configuration, the system reaches a legitimate configuration within $\mathcal{O}(n)$ rounds, w.h.p.*

The proof relies on the analysis of the behaviour of some essential random variables describing the repeated balls-into-bins process. In the next paragraph, we informally describe the main steps of this analysis. Then in sections 7.1.1-7.1.3, we prove the technical results required by such steps and, finally, in Section 7.1.4 these technical results are combined in order to prove Theorem 12.

[1] As usual in this work, we use capital letters for random variables, lower case for quantities, and bold for vectors.

7.1.0.1. *Overview of the analysis.* In the repeated balls-into-bins process, every bin can release at most one ball per round. As a consequence, the random walks performed by the balls delay each other and are thus correlated in a way that can make bin queues larger than in the independent case. Indeed, intuitively speaking, a large load observed at a bin in some round makes "any" ball more likely to spend several future rounds in that bin, because if the ball ends up in that bin in one of the next few rounds, it undergoes a large delay. This is essentially the major technical issue to cope with.

The previous approach in Section 6.3 relies on the fact that, in every round, the expected balance between the number of incoming and outgoing balls is always non-positive for every non-empty bin (notice that the expected number of incoming balls is always at most one). This may suggest viewing the process as a sort of parallel *birth-death* process [**LPW09**]. Using this approach and with some further arguments, one can (only) get the "standard-deviation" bound $\mathcal{O}(\sqrt{t})$ in Section 6.3.

The analysis presented here, which proves Theorem 12, proceeds along three main steps.

i) We first show that, after the first round, the aforementioned expected balance is always negative, namely, not larger than $-1/4$. Indeed, the number of empty bins remains at least $n/4$ with (very) high probability, which is extremely useful since a bin can only receive tokens from non-empty bins. This fact is shown to hold starting from *any* configuration and over any period of polynomial length.

ii) In order to exploit the above negative balance to bound the load of the bins, we need some strong concentration bound on the number of balls entering a specific bin u along any period of polynomial size. However, it is easy to see that, for any fixed u, the random variables $\{Z_u^{(t)}\}_{t\geqslant 0}$ counting the number of balls entering bin u are not mutually independent, neither are they negatively associated, so that we cannot apply standard tools to prove concentration, as we show in Section 7.2. To address this issue, we define a simpler repeated balls-into-bins process as follows.

TETRIS PROCESS.
Starting from any configuration with at least $n/4$ empty bins, in each round:

- from every non-empty bin we pick one ball and we throw it away, and
- we pick exactly $(3/4)n$ *new balls* and we put each of them independently and u.a.r. in one of the n bins.

Using a coupling argument and our previous upper bound on the number of empty bins, we prove that the maximum number of balls accumulating in a bin in the original process is not larger than the maximum number of balls accumulating in a bin in the TETRIS process, w.h.p.

iii) The TETRIS process is simpler than the original one since, at every round, the number of balls assigned to the bins does not depend on the system's state in the previous round. Hence, random variables $\{\hat{Z}_u^{(t)}\}_{t\geqslant 0}$ counting the number of balls arriving at bin u in the TETRIS process are mutually independent. We can thus apply standard concentration bounds. On the other hand, differently from the approximating process considered in Section 6.3, the negative balance of incoming and outgoing balls proved in Step i) still holds, thus yielding a much smaller bound on the maximum load than that in Section 6.3.

In the remainder of this section, we formally describe the above three steps. Lastly, we prove Theorem 12 (in Section 7.1.4).

7.1.1. On the number of empty bins

We next show that the number of *empty* bins is at least a constant fraction of n over a very large time-window, w.h.p. This fact could be proved by standard concentration arguments if, at every round, *all* balls were thrown independently and uniformly at random. A little care is instead required in our process to properly handle, at any round, "congested" bins whose load exceeds 1. These bins are surely non-empty at the next round too. So, the number of empty bins at a given round also depends on the number of congested bins in the previous round.

LEMMA 48. *Let $\mathbf{q} = (q_1, \ldots, q_n)$ be a configuration in a given round and let X be the random variable indicating the number of empty bins in the next round. For any large enough n, it holds that*

$$\Pr\left(X \leqslant \frac{n}{4}\right) \leqslant e^{-\alpha n},$$

where α is a suitable positive constant.

PROOF. Let $a = a(\mathbf{q})$ and $b = b(\mathbf{q})$ respectively denote the number of empty bins and the number of bins with exactly one token in configuration $\mathbf{q}$. For each bin u of the $a + b$ bins with at most one token, let Y_u be the random variable indicating whether or not bin u is empty in the next round, so that

$$X = \sum_{u=1}^{a+b} Y_u \quad \text{and} \quad \Pr\left(Y_u = 1\right) = \left(1 - \frac{1}{n}\right)^{n-a} \geqslant e^{-\frac{n-a}{n-1}},$$

where in the last inequality we used the fact that $1 - x \geqslant e^{-\frac{x}{1-x}}$. Hence we have that

$$\mathbb{E}\left[X\right] \geqslant (a+b)\, e^{-\frac{n-a}{n-1}}. \tag{139}$$

The crucial fact is that the number of bins with two or more tokens cannot exceed the number of empty bins, i.e.

$$n - (a+b) \leqslant a.$$

Thus, we can bound the number of empty bins from below[2], $a \geqslant (n-b)/2$, and by using that bound in (139) we get

$$\mathbb{E}[X] \geqslant \frac{n+b}{2}\, e^{-\frac{n+b}{2(n-1)}}.$$

Now observe that, for large enough n a positive constant ε exists such that

$$\frac{n+b}{2}\, e^{-\frac{n+b}{2(n-1)}} \geqslant (1+\varepsilon)\frac{n}{4},$$

for every $0 \leqslant b \leqslant n$.

As a consequence of propositions 7 and 11 in [**DR98**], it follows that the random variables $Y_1, \ldots, Y_{a+b}$ are *negatively associated* (Definition 18). Thus we can apply (see Lemma 7 in [**DR98**]) the Chernoff bound eqrefCB:lowertail in Lemma 76 with $\delta = \varepsilon/(1+\varepsilon))$ to r.v. X to obtain

$$\Pr\left(X \leqslant \frac{n}{4}\right) \leqslant \exp\left(-\frac{\varepsilon^2}{4(1+\varepsilon)}n\right).$$

gg

From the above lemma it follows that, if we look at our process over a time-window $T = T(n)$ of polynomial size, after the first round we always see at least $n/4$ empty bins, w.h.p. More formally, for every $t \in \{1, \ldots, T\}$, let $\mathcal{E}_t$ be the event "The number of empty bins at round t is at least $n/4$". From Lemma 51 and the union bound we get the following lemma.

LEMMA 49. *Let $\mathbf{q}_0$ denote the initial configuration, let $T = T(n) = n^c$ for an arbitrarily large constant c. For any large enough n it holds that*

$$\Pr\left(\bigcap_{t=1}^{T} \mathcal{E}_t \mid \mathbf{Q}^{(0)} = \mathbf{q}_0\right) \geqslant 1 - e^{-\gamma n},$$

where γ is a suitable positive constant.

PROOF. By using the union bound we have that

$$\begin{aligned}\Pr\left(\bigcap_{t=1}^{T} \mathcal{E}_t \mid \mathbf{Q}^{(0)} = \mathbf{q}_0\right) &= 1 - \Pr\left(\bigcup_{t=1}^{T} \overline{\mathcal{E}_t} \mid \mathbf{Q}^{(0)} = \mathbf{q}_0\right) \\ &\geqslant 1 - \sum_{t=1}^{T} \Pr\left(\overline{\mathcal{E}_t} \mid \mathbf{Q}^{(0)} = \mathbf{q}_0\right).\end{aligned}$$

By conditioning on the configuration at round $t-1$, from the Markov property and Lemma 48 it then follows that

$$\begin{aligned}&\Pr\left(\overline{\mathcal{E}_t} \mid \mathbf{Q}^{(0)} = \mathbf{q}_0\right) \\ &= \sum_{\mathbf{q}} \Pr\left(\overline{\mathcal{E}_t} \mid \mathbf{Q}^{(t-1)} = \mathbf{q}\right) \Pr\left(\mathbf{Q}^{(t-1)} = \mathbf{q} \mid \mathbf{Q}^{(0)} = \mathbf{q}_0\right) \leqslant e^{-\alpha n}.\end{aligned}$$

[2]Observe that this argument only works to get a *lower* bound on the number of empty bins and not for an upper bound.

Hence,

$$\Pr\left(\bigcap_{t=1}^{T}\mathcal{E}_t \mid \mathbf{Q}^{(0)}=\mathbf{q}_0\right)\geqslant 1-Te^{-\alpha n}\geqslant 1-e^{-\gamma n},$$

for a suitable positive constant γ. ∎

7.1.2. Coupling with Tetris

Using a coupling argument and Lemma 49 we now prove that the maximum load in the original process is stochastically not larger than the maximum load in the Tetris process, w.h.p.

In what follows we denote by $W^{(t)}$ the *set* of non-empty bins at round t in the original process. Recall that, in the latter, at every round a ball is selected from every non-empty bin u and it is moved to a bin chosen u.a.r. Accordingly we define, for every round t, the random variables

$$\left\{X_u^{(t+1)}\,:\,u\in W^{(t)}\right\},$$

where $X_u^{(t+1)}$ indicates the new position reached in round $t+1$ by the ball selected in round t from bin u. Notice that for every non-empty bin $u\in W^{(t)}$ we have that

$$\Pr\left(X_u^{(t+1)}=v\right)=\frac{1}{n},$$

for every bin $v\in[n]$. The random process $\left\{\mathbf{Q}^{(t)}\,:\,t\in\mathbb{N}\right\}$ is completely defined by random variables X_u^t's, indeed we can write

$$\mathcal{Q}_v^{(t+1)}=\mathcal{Q}_v^{(t)}\dot{-}\,1+\left|\left\{u\in W^{(t)}\,:\,X_u^{(t+1)}=v\right\}\right|,$$

and

$$W^{(t+1)}=\left\{u\in[n]\,:\,\mathcal{Q}_u^{(t+1)}\geqslant 1\right\},$$

where we used notation $a\dot{-}\,b=\max\{a-b,0\}$. Analogously, for each bin $u\in[n]$ in the Tetris process, let $\hat{\mathcal{Q}}_u^{(t)}$ be the random variable indicating the number of balls in bin u in round t. We next prove that, over any polynomially-large time window, the maximum load of any bin in our process is stochastically smaller than the maximum number of balls in a bin of the Tetris process, w.h.p. More formally, we prove the following lemma.

Lemma 50. *Assume we start our process and the* Tetris *process from the same initial configuration* $\mathbf{q}=(q_1,\ldots,q_n)$ *such that* $\sum_{u=1}^{n}q_u=n$ *and containing at least* $n/4$ *empty bins. Let* $T=T(n)$ *be an arbitrary round and let* M_T *and* $\hat{M}_T$ *be respectively the random variables indicating the maximum loads in our original process and in the* Tetris *process, up to round* T*. Formally*

$$M_T=\max\{\mathcal{Q}_u^{(t)}\,:\,u\in[n],\,t=1,2,\ldots,T\},$$
$$\hat{M}_T=\max\{\hat{\mathcal{Q}}_u^{(t)}\,:\,u\in[n],\,t=1,2,\ldots,T\}.$$

For every $k \geqslant 0$ it holds that

$$\Pr\left(M_T \geqslant k\right) \leqslant \Pr\left(\hat{M}_T \geqslant k\right) + T \cdot e^{-\gamma n},$$

for a suitable positive constant γ.

PROOF. We proceed by coupling the TETRIS process with the original one round by round. Intuitively speaking the coupling proceeds as follows:

- **Case (i).** *the number of non-empty bins in the original process is $k \leqslant \frac{3}{4}n$.* For each non-empty bin u, let i_u be the ball picked from u. We throw one of the $\frac{3}{4}n$ new balls of the TETRIS process in the same bin in which i_u ends up. Then, we throw all the remaining $\frac{3}{4}n - k$ balls independently u.a.r.
- **Case (ii).** *the number of non-empty bins is $k > \frac{3}{4}n$.* We run one round of the TETRIS process independently from the original one.

By construction, if the number of non-empty bins in the original process is not larger than $\frac{3}{4}n$ at any round, Case (ii) never applies and the TETRIS process "dominates" the original one, meaning that every bin in the TETRIS process contains at least as many balls as the corresponding bin in the original one. Since from Lemma 49 we know that the number of non-empty bins in the original process is not larger than $\frac{3}{4}n$ for any time-window of polynomial size, w.h.p., we thus have that the TETRIS process dominates the original process for the whole time window, w.h.p.

More formally, for $t \in \{1, \ldots, T\}$, denote by $B^{(t)}$ the set of new balls in the TETRIS process at round t (recall that the size of $B^{(t)}$ is $(3/4)n$ for every $t \in \{1, \ldots, T\}$). For any round t and any ball $i \in B^{(t)}$, let $\hat{X}_i^{(t)}$ be the random variable indicating the bin where the ball ends up. Finally, let

$$\left\{U_i^{(t)} \ : \ t = 1, \ldots, T,\, i \in B^{(t)}\right\}$$

be a family of i.i.d. random variables uniform over $[n]$. At any round $t \in \{1, \ldots, T\}$, we have to distinguish two cases:

- **Case $|W^{(t-1)}| \leqslant (3/4)n$.** Let $B_W^{(t)}$ be an arbitrary subset of $B^{(t)}$ with size exactly $|W^{(t-1)}|$, let $f^{(t)} \ : \ B_W^{(t)} \to W^{(t-1)}$ be an arbitrary bijection and set

$$\hat{X}_i^{(t)} = \begin{cases} X_i^{(t)} & \text{if } i \in B_W^{(t)}, \\ U_i^{(t)} & \text{if } i \in B^{(t)} \setminus B_W^{(t)}. \end{cases}$$

- **Case $|W^{(t-1)}| > (3/4)n$.** Set $\hat{X}_i^{(t)} = U_i^{(t)}$ for all $i \in B^{(t)}$.

By construction we have that random variables

$$\left\{\hat{X}_i^{(t)} \ : \ t \in \{1, 2, \ldots, T\},\, i \in B^{(t)}\right\}$$

are mutually independent and uniformly distributed over $[n]$. Moreover, in the joint probability space for any k we have that

$$\begin{aligned}\Pr\left(M_T \geqslant k\right) &= \Pr\left(M_T \geqslant k,\, \hat{M}_T \geqslant M_t\right) + \Pr\left(M_T \geqslant k,\, \hat{M}_T < M_T\right) \\ &\leqslant \Pr\left(\hat{M}_T \geqslant k\right) + \Pr\left(\hat{M}_T < M_T\right).\end{aligned}$$

Finally, let $\mathcal{E}_T$ be the event "There are at least $n/4$ empty bins at all rounds $t \in \{1, \ldots, T\}$" and observe that, from the coupling we have defined, the event $\mathcal{E}_T$ implies event "$\hat{M}_T \geqslant M_T$". Hence

$$\Pr\left(\hat{M}_T < M_T\right) \leqslant \Pr\left(\overline{\mathcal{E}_T}\right),$$

and the thesis follows from Lemma 49. ∎

7.1.3. Analysis of the Tetris process

We begin by observing that in the Tetris process, the random variables indicating the number of balls ending up in a bin in different rounds are i.i.d. binomial. This fact is extremely useful to give upper bounds on the load of the bins, as we do in the next simple lemma, that we use to prove self-stabilization of the original process.

Lemma 51. *From any initial configuration, in the* Tetris *process every bin is empty at least once within* $5n$ *rounds, w.h.p.*

Proof. Let $u \in [n]$ be a bin with $k \leqslant n$ balls in the initial configuration. For $t \in \{1, \ldots, 5n\}$ let Y_t be the random variable indicating the number of new balls ending up in bin u at round t. Notice that in the Tetris process $Y_1, \ldots, Y_{5n}$ are i.i.d. $B\left((3/4)n,\, 1/n\right)$ hence

$$\mathbb{E}\left[Y_1 + \cdots + Y_{5n}\right] = (15/4)n,$$

and by applying Chernoff bound (Lemma 76) with $\delta = 1/15$ we get

$$\Pr\left(Y_1 + \cdots + Y_{5n} \geqslant 4n\right) \leqslant e^{-\alpha n},$$

where $\alpha = 1/(180)$.

Now let $\mathcal{E}_u$ be the event *"Bin u is non-empty for all the $5n$ rounds"*. Since when a bin is non-empty it looses a ball at every round, event $\mathcal{E}_u$ implies, in particular, that

$$k - 5n + Y_1 + \cdots + Y_{5n} \geqslant 0,$$

that is

$$Y_1 + \cdots + Y_{5n} \geqslant 5n - k \geqslant 4n.$$

Thus

$$\Pr\left(\mathcal{E}_u\right) \leqslant \Pr\left(Y_1 + \cdots + Y_{5n} \geqslant 4n\right) \leqslant e^{-\alpha n}.$$

The thesis follows from the union bound over all bins $u \in [n]$. ∎

We next focus on the maximum load that can be observed in the TETRIS process at any given bin within a finite interval of time. We note that this result could be proved using tools from *drift analysis* (e.g., see [**Haj82**]). We provide here an elementary and direct proof, that explicitly relies on the Markovian structure of the TETRIS process.

Let $\{X_t\}_t$ be a sequence of i.i.d. $B\left((3/4)n, 1/n\right)$ random variables and let Z_t be the Markov chain with state space $\{0, 1, 2, \dots\}$ defined as follows

$$Z_t = \begin{cases} 0 & \text{if } Z_{t-1} = 0, \\ Z_{t-1} - 1 + X_t & \text{if } Z_{t-1} \geqslant 1. \end{cases} \tag{140}$$

Observe that 0 is an absorbing state for Z_t and let τ be the absorption time

$$\tau = \inf\{t \in \mathbb{N} \,:\, Z_t = 0\}.$$

We first prove the following lemma.

LEMMA 52. *For any initial starting state $k \in \mathbb{N}$ and any $t \geqslant 8k$, it holds that*

$$\Pr\left(\tau > t \,|\, Z_0 = k\right) \leqslant e^{-t/144}.$$

PROOF. Observe that

$$\begin{aligned} \Pr\left(\tau > t \,|\, Z_0 = k\right) &= \Pr\left(Z_t > 0 \,|\, Z_0 = k\right) \\ &= \Pr\left(k + \sum_{i=1}^{t} X_i - t > 0\right) \\ &= \Pr\left(\sum_{i=1}^{t} X_i > t - k\right) \leqslant \Pr\left(\sum_{i=1}^{t} X_i > \frac{7}{8}t\right), \end{aligned}$$

where in the last inequality we used hypothesis $k < (1/8)t$. Since the X_is are i.i.d. binomial $B((3/4)n, 1/n)$, it follows that $\sum_{i=1}^{t} X_i$ is binomial $B((3/4)nt, 1/n)$ and from Chernoff bound (Lemma 25) we have that

$$\begin{aligned} \Pr\left(\sum_{i=1}^{t} X_i > \frac{7}{8}t\right) &= \Pr\left(\sum_{i=1}^{t} X_i > \left(1 + \frac{1}{6}\right)\frac{3}{4}t\right) \\ &\leqslant e^{-\frac{(1/6)^2}{3}\frac{3}{4}t} = e^{-t/144}. \end{aligned}$$

gg

Now we can easily prove the following statement on the TETRIS process.

LEMMA 53. *Let c be an arbitrarily-large constant, and let the* TETRIS *process start from any legitimate configuration. The maximum load $\hat{M}^{(t)}$ is $\mathcal{O}(\log n)$ for all $t = \mathcal{O}(n^c)$, w.h.p.*

PROOF. Consider an arbitrary bin u that is non-empty in the initial legitimate configuration. Let $\hat{\mathcal{Q}}^{(0)} = \mathcal{O}(\log n)$ be its initial load[3] and let

$$\tau = \inf\left\{t \,:\, \hat{\mathcal{Q}}^{(t)} = 0\right\}$$

be the first round the bin becomes empty. Observe that, for any $t \leqslant \tau$, $\hat{\mathcal{Q}}^{(t)}$ behaves exactly as the Markov chain defined in (140). Hence, from Lemma 52 it follows that for every constant $\hat{c}$ such that $\hat{c}\log n \geqslant 8\hat{\mathcal{Q}}^{(0)}$ we have

$$\Pr(\tau > \hat{c}\log n) \leqslant n^{-\frac{\hat{c}}{144}}. \tag{141}$$

Thus, within $\mathcal{O}(\log n)$ rounds the bin is empty, w.h.p., and since the load of the bin decreases of at most one unit per round, the load of the bin is $\mathcal{O}(\log n)$ for all such rounds, w.h.p.

Next, define a *phase* as any sequence of rounds that starts when the bin becomes non-empty and ends when it becomes empty again. Notice that, by using a standard balls-into-bins argument, in the first round of each phase the load of the bin is $\mathcal{O}(\log n/\log\log n)$, w.h.p. Moreover, in any phase the load of the bin can be coupled with the Markov chain in (140). Hence, for any arbitrary large constant c we can choose the constant $\hat{c}$ in (141) large enough so that, by taking the union bound over all phases up to round n^c, the load of the bin is $\mathcal{O}(\log n)$ in all rounds $t \leqslant n^c$, w.h.p.

Finally, observe that for any bin that is initially empty the same argument applies with the only difference that the first phase for the bin does not start at round 0 but at the first round the bin becomes non-empty. The thesis thus follows from a union bound over all the bins. ∎

7.1.4. Back to the original process

We are now ready to prove the main theorem of the chapter.

THEOREM 12 (Repeated Balls into Bins Max Load). *Let c be an arbitrarily-large constant and let $\mathbf{q}$ be any legitimate configuration. Let the* repeated balls-into-bins process *start from $\mathbf{Q}^{(0)} = \mathbf{q}$. Then, over any period of length $\mathcal{O}(n^c)$, the process visits only legitimate configurations, w.h.p., i.e. $M^{(t)} = \mathcal{O}(\log n)$ for all $t = \mathcal{O}(n^c)$, w.h.p. Moreover, starting from any configuration, the system reaches a legitimate configuration within $\mathcal{O}(n)$ rounds, w.h.p.*

PROOF OF THEOREM 12. From a standard balls-into-bins argument (see, e.g., [**MU05**]), starting from any legitimate configuration, after one round the process still lies in a legitimate configuration, w.h.p. Moreover, thanks to Lemma 48, there are at least $n/4$ empty bins, w.h.p. From Lemma 50 with $T = \mathcal{O}\left(n^c\right)$, we have that the maximum load of the repeated balls-into-bins process does not exceed the maximum load of the TETRIS process in all rounds $1, \ldots, T$, w.h.p. Finally, the upper bound on the maximum load of the TETRIS process in Lemma 53 completes the proof of the first statement of Theorem 12.

[3]We omit the subscript u in the remainder of this proof since clear from context.

As for self-stabilization, given an arbitrary initial configuration, Lemma 51 implies that within $\mathcal{O}(n)$ rounds, all bins have been emptied at least once, w.h.p. When a bin becomes empty, Lemma 52 ensures that its load is $\mathcal{O}(\log n)$ over a polynomial number of rounds. Hence, within $\mathcal{O}(n)$ rounds, the system reaches a legitimate configuration, w.h.p. gs

7.2. Negative Association

In this section we give a simple counterexample showing that, in our balls-into-bins process, the random variables counting the number of balls arriving in a given bin in different rounds cannot be negatively associated. We first recall the definition of negative association.

DEFINITION 18 (Negative association). Random variables $X_1, \ldots, X_n$ are *negatively associated* if, for every pair of disjoint subsets $I, J \subseteq [n]$, it holds that

$$\mathbb{E}\left[f\left(X_i,\, i \in I\right) \cdot g\left(X_j,\, j \in J\right)\right] \leqslant \mathbb{E}\left[f\left(X_i,\, i \in I\right)\right] \cdot \mathbb{E}\left[g\left(X_j,\, j \in J\right)\right]$$

for all pairs of functions $f : \mathbb{R}^{|I|} \to \mathbb{R}$ and $g : \mathbb{R}^{|J|} \to \mathbb{R}$ that are both non-decreasing or both non-increasing.

Consider our random process with $n = 2$ and let X_1 and X_2 be the random variables indicating the number of tokens arriving at the first bin in rounds 1 and 2, respectively. Let $f \equiv g$ be the non-increasing function

$$f(x) = \begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{if } x > 0. \end{cases}$$

If X_1 and X_2 were negatively associated, we would have that

$$\Pr(X_1 = 0,\, X_2 = 0) \leqslant \Pr(X_1 = 0)\Pr(X_2 = 0).$$

However, by direct calculation it is easy to compute that

$$\Pr(X_1 = 0,\, X_2 = 0) = \frac{1}{8},$$

because, in order for "$X_1 = 0,\, X_2 = 0$" to happen, at the first round both balls have to end up in the second bin (this happens with probability $1/4$) and at the second round the ball chosen in the second bin has to stay there (this happens with probability $1/2$). We also have that $\Pr(X_1 = 0) = 1/4$ and by conditioning on all the three possible configurations at round 1 we have $\Pr(X_2 = 0) = 3/8$. Thus

$$\frac{1}{8} = \Pr(X_1 = 0,\, X_2 = 0) > \Pr(X_1 = 0)\Pr(X_2 = 0) = \frac{1}{4} \cdot \frac{3}{8}.$$

In general, intuitively speaking it seems that event "$X_t = 0$" makes more likely the event that there are a lot of empty bins in the system, which in turn makes more likely event "$X_{t+1} = 0$" that the bin receives no tokens at round $t + 1$ as well.

7.3. Parallel Resource Assignment

In this section, we resume the original interpretation of the repeated balls-into-bins process as running parallel random walks of n distinct tokens (i.e. balls), each of them starting from a node (i.e. bins) of the complete graph of size n. This is a randomized protocol for the parallel allocation problem where tokens represent different resources/tasks that must be assigned to all nodes in mutual exclusion [**Coo11**]. In this scenario, a critical complexity measure is the (global) cover time, i.e., the time required by any token to visit all nodes.

It is important to observe that our analysis of the maximum load works for anonymous tokens and nodes and, hence, for any particular queuing strategy. Under FIFO strategy, no token spends in a bin a number of rounds exceeding the current load as it entered the bin. Theorem 12 then implies that, after an initial stabilizing phase of $\mathcal{O}(n)$ rounds, every token spends at most a logarithmic number of rounds in any bin queue it traverses and over any period of polynomial length, w.h.p. We also know that the cover time of the single random-walk process is $\mathcal{O}(n \log n)$, w.h.p. (see, e.g., [**MU05**]). Combining the above two facts, we get the following, almost tight result on the Parallel Resource Assignment problem.

Corollary 8 (Parallel Resource Assignment). *The random-walk protocol for the Parallel Resource Assignment problem on the clique has cover time* $\mathcal{O}\left(n \log^2 n\right)$*, w.h.p.*

7.3.0.1. *Adversarial model.* The self-stabilization property shown in Theorem 12 makes the random walk protocol robust to transient faults. We can consider an adversarial model in which, in some *faulty* rounds, an adversary can reassign the tokens to the nodes in an arbitrary way. Then, the linear convergence time shown in Theorem 12 implies that the $\mathcal{O}\left(n \log^2 n\right)$ bound on the cover time still holds provided the faulty rounds happen with a frequency not higher than γn, for any constant $\gamma \geqslant 6$. Indeed, thanks to Lemma 51, the action of an adversary manipulating the system configuration once every γn rounds can affect only the successive $5n$ rounds, while our analysis in the non-adversarial model does hold for the remaining $(\gamma - 5)n$ rounds. It follows that the overall slowdown on the cover time produced by such an adversary is at most a constant factor on the previous $\mathcal{O}\left(n \log^2 n\right)$ upper bound, w.h.p.

CHAPTER 8

Consensus Despite Noise

In this Chapter we prove the results discussed in Section 2.5.

While error-correcting codes are efficient methods for handling *noisy* communication channels in the context of technological networks, such elaborate methods differ a lot from the unsophisticated way biological entities are supposed to communicate. Yet, in [**FKP11**] it has been shown that complex coordination tasks such as *bit dissemination* and *majority consensus* can plausibly be achieved in biological systems subject to noisy communication channels, where every message transferred through a channel remains intact with small probability $\frac{1}{2} + \varepsilon$, without using coding techniques. The previous result is a considerable step towards a better understanding of the way biological entities may cooperate. It has nevertheless been established only in the case of 2-valued *opinions*: rumor spreading aims at broadcasting a single-bit opinion to all nodes, and majority consensus aims at leading all nodes to adopt the single-bit opinion that was initially present in the system with (relative) majority. In this chapter, we extend this previous work to k-valued opinions, for any constant $k \geqslant 2$. This extension requires to address a series of important issues, some conceptual, others technical. We have to revisit entirely the notion of noise, for handling channels carrying *k-valued* messages. In fact, we precisely characterize the type of noise patterns for which plurality consensus is solvable. Also, a key result employed in the bivalued case by Feinerman et al. is an estimate of the probability of observing the most frequent opinion from observing the mode of a small sample. We generalize this result to the multivalued case by providing a new analytical proof for the bivalued case that is amenable to be extended, by induction, and that is of independent interest.

8.1. Model and Results in the Noisy Setting

In this section we formally define the communication model, the main definitions, the investigated problems and the results that we prove, part of which has already been introduced in Section 2.5.

We do not provide a definition of what is a biologically feasible protocol, since the computational investigation with this respect is still too premature for such an attempt. Nevertheless, we remark that intuitively we look for protocols that, if not dynamics, are at least simple enough to be plausible communication strategies for primitive biological system. As the reader can see in sections 8.1.3 and 8.2.1, we consider a natural generalization of the

protocol given in [**FHK15**], which is plainly an elementary combination of sampling and majority operations.

8.1.1. Communication model and definition of the problems

We consider the same communication model considered in Chapter 7, the *uniform $\mathcal{PUSH}$ model* [**DGH$^+$87**], where in each (synchronous) round each agent can send *(push)* a message to another agent chosen uniformly at random. This occurs without having the sender or the receiver learning about each other's identity. Note that it may happen that several agents push a message to the same node u at the same round. In the latter case we assume that the nodes receive them in a random order; for a detailed discussion regarding this assumption, we refer the reader to Section 8.4.

We study the problems of bit dissemination and plurality consensus. In both cases, we assume that nodes can support opinions represented by an integer in $[k] = \{1, \ldots, k\}$. Additionally, there may be *undecided* nodes that do not support any opinion, which represents nodes that are not actively aware that the system has started to solve the problem; thus, undecided nodes are not allowed to send any message before receiving any of them.

- In bit dissemination, initially, one node, called the source, has an opinion $m \in \{1, \ldots, k\}$, called the *correct opinion*. All the other nodes have no opinion. The objective is to design a protocol insuring that, after a certain number of communication rounds, every node has the correct opinion m.
- In plurality consensus, initially, for every $i \in \{1, \ldots, k\}$, a set A_i of nodes have opinion i. The sets A_i, $i = 1, \ldots, k$, are pairwise disjoint, and their union does not need to cover all nodes, i.e., there may be some *undecided* nodes with no opinion initially. The objective is to design a protocol insuring that, after a certain number of communication rounds, every node has the plurality opinion, that is, the opinion m with relative majority in the initial setting (i.e., $|A_m| > |A_j|$ for any $j \neq m$).

Observe that the bit dissemination problem is a special case of the plurality consensus problem with $|A_m| = 1$ and $|A_j| = 0$ for any $j \neq m$.

Following the guidelines of [**FHK14**], we work under two constraints:

(1) We restrict ourselves to protocols in which each node can only transmit opinions, i.e., every message is an integer in $\{1, \ldots, k\}$.
(2) Transmissions are subject to noise, that is, for every round, and for every node u, if an opinion $i \in \{1, \ldots, k\}$ is transmitted to node u during that round, then node u will receive message $j \in \{1, \ldots, k\}$ with probability $p_{i,j} \geqslant 0$, where $\sum_{j=1}^{k} p_{i,j} = 1$.

The *noisy push model* is the uniform $\mathcal{PUSH}$ model together with the previous two constraints. The probabilities $\{p_{i,j}\}_{i,j\in[k]}$ can be seen as a transition matrix, called the *noise matrix*, and denoted by $P = (p_{i,j})_{i,j\in[k]}$ (see Figure

25). The noise matrix in [**FHK14**] is simply

$$P = \begin{pmatrix} \frac{1}{2}+\varepsilon & \frac{1}{2}-\varepsilon \\ \frac{1}{2}-\varepsilon & \frac{1}{2}+\varepsilon \end{pmatrix}. \tag{142}$$

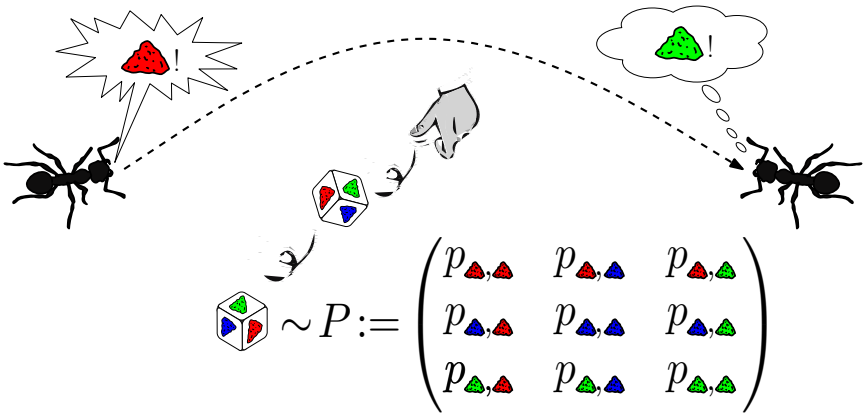

FIGURE 25. A representation of the action of the noise. After an agent sends message i, a k-sided die specific for message i is thrown. The received message is determined by the uppermost face of the die when it comes to rest. The die end up on face j with probability $p_{i,j}$, defined by the noise matrix P.

8.1.2. Plurality bias, and majority preservation

When time proceeds, our protocol results in the proportion of nodes with a given opinion to evolve. Note that there might be nodes who do not support any opinion at time t. As mentioned in the previous section, we call such nodes *undecided*. We denote by $a^{(t)}$ the fraction of nodes supporting any opinion at time t and we call the nodes contributing to $a^{(t)}$ *decided*. Observe that the fraction of undecided nodes at time t is then $1 - a^{(t)}$. Let $c_i^{(t)}$ be the fraction of decided nodes in the system that support opinion $i \in [k]$ at the beginning of round t, so that $\sum_{i\in[k]} c_i^{(t)} = a^{(t)}$. We remark the different meaning of the latter notation compared to that used in chapters 5 and 6. Let $\hat{c}_i^{(t)}$ be the fraction of decided nodes which receive at least one message at time $t-1$ and support opinion $i \in [k]$ at the beginning of round t. We write $\mathbf{c}^{(t)} = (c_1^{(t)}, ..., c_k^{(t)})$ to denote the *opinion distribution* of the opinions at time t. Similarly, let $\hat{\mathbf{c}}^{(t)} = (\hat{c}_1^{(t)}, ..., \hat{c}_k^{(t)})$. In particular, if every node would

simply switch to the last opinion it received, then

$$\mathbb{E}[\hat{c}_i^{(t+1)} \mid \mathbf{c}^{(t)}] = \sum_{j\in[k]} \Pr[\text{received } i \mid \text{original message is } j] \cdot \Pr[\text{original message is } j] = \sum_{j\in[k]} c_j^{(t)} \cdot p_{j,i}.$$

That is,

$$\mathbb{E}[\hat{\mathbf{c}}^{(t+1)} \mid \mathbf{c}^{(t)}] = \mathbf{c}^{(t)} \cdot P, \tag{143}$$

where P is the noise matrix. In particular, in the absence of noise, we have $P = I$ (the identity matrix), and if every node would simply copy the opinion that it just received, we had $\mathbb{E}[\hat{\mathbf{c}}^{(t+1)} \mid \mathbf{c}^{(t)}] = \mathbf{c}^{(t)}$. So, given the opinion distribution at round t, from the definition of the model it follows that the messages each node receives at round $t+1$ can equivalently be seen as being sent from a system without noise, but whose opinion distribution at round t is $\mathbf{c}^{(t)} \cdot P$.

Recall that m denotes the initially correct opinion, that is, the source's opinion in the bit dissemination problem, and the initial plurality opinion in the plurality consensus problem. The following definition naturally extends the concept of *majority bias* in [**FHK14**] to *plurality bias*.

DEFINITION 19 (δ-Biased Configuration). *Let $\delta > 0$. An opinion distribution $\mathbf{c}$ is said to be* δ-biased toward opinion m *if $c_m - c_i \geqslant \delta$ for all $i \neq m$.*

In [**FHK14**], each binary opinion that is transmitted between two nodes is flipped with probability at most $\frac{1}{2} - \varepsilon$, with[1] $\varepsilon = n^{-\frac{1}{4}+\eta}$ for an arbitrarily small $\eta > 0$. Thus, the noise is parametrized by ε. The smaller ε, the more noisy are the communications. We generalize the role of this parameter with the following definition.

DEFINITION 20 ((ε,δ)-m.p. Noise Matrix). *Let $\varepsilon = \varepsilon(n)$ and $\delta = \delta(n)$ be positive. A noise matrix P is said to be* (ε,δ)-majority-preserving ((ε,δ)-m.p.) with respect to opinion m *if, for every opinion distribution $\mathbf{c}$ that is δ-biased toward opinion m, we have*

$$(\mathbf{c} \cdot P)_m - (\mathbf{c} \cdot P)_i > \varepsilon\,\delta$$

for all $i \neq m$.

In the bit dissemination problem, as well as in the plurality consensus problem, when we say that a noise matrix is (ε,δ)-m.p., we implicitly mean that it is (ε,δ)-m.p. with respect to the initially correct opinion. We defer a discussion on the class of (ε,δ)-m.p. noise matrices in Section 8.3 (including the tightness of the class w.r.t. theorems 13 and 14).

[1] For a discussion on what happens for other values of ε, see Section 8.6.

8.1.3. Formal statements of the results

We show that a natural generalization of the protocol in [**FHK14**] solves the bit dissemination problem and the plurality consensus problem for an arbitrary number of opinions k. More precisely, using the protocol which we describe in Section 8.2.1, we can establish the following two results, whose proof can be found in Section 8.2.

THEOREM 13 (Noisy Bit Dissemination). *Assume that the noise matrix P is (ε,δ)-m.p. with $\varepsilon = \Omega(n^{-\frac{1}{4}+\eta})$ for an arbitrarily small constant $\eta > 0$ and $\delta = \Omega(\sqrt{\log n/n})$. There exists a protocol, using $O(\log\log n + \log\frac{1}{\varepsilon})$ bits of memory at each node, which solves the noisy bit dissemination problem with k opinions in $O(\frac{\log n}{\varepsilon^2})$ communication rounds, w.h.p.*

THEOREM 14 (Noisy Plurality Consensus). *Let S with $|S| = \Omega(\frac{1}{\varepsilon^2}\log n)$ be an initial set of nodes with opinions in $[k]$, the rest of the nodes having no opinions. Assume that the noise matrix P is (ε,δ)-m.p. for some $\varepsilon > 0$, and that S is $\Omega(\sqrt{\log n/|S|})$-majority-biased. There exists a protocol, using $O(\log\log n + \log\frac{1}{\varepsilon})$ bits of memory at each node, which solves the noisy plurality consensus problem with k opinions in $O(\frac{\log n}{\varepsilon^2})$ communication rounds, w.h.p.*

For $k = 2$, we get the theorems in [**FHK14**] from the above two theorems. Indeed, the simple 2-dimensional noise matrix of (142) is ε-majority-biased. Note that, as in [**FHK14**], the plurality consensus algorithm requires the nodes to know the size $|S|$ of the set S of decided nodes.

8.2. The Analysis

In this section we prove Theorem 13 and Theorem 14 by presenting a more general analysis of Stage 1 than that given in [**FHK14**] and a new analysis of Stage 2. Note that the proof techniques required for the generalization to arbitrary k significantly depart from those in [**FHK14**] for the case $k = 2$. In particular, our approach provides a general framework for rigorously dealing with many kinds of stochastic dependences among messages in the uniform $\mathcal{PUSH}$ model.

8.2.1. Definition of the Protocol

We describe a bit dissemination protocol performing in two *stages*. Each stage is decomposed into a number of *phases*, each one decomposed into a number of *rounds*. During each phase of the two stages, the nodes apply the simple rules given below.

8.2.1.1. *The rule during each phase of Stage 1.* Nodes that already support some opinion at the beginning of the phase push their opinion at each round of the phase. Nodes that do not support any opinion at the beginning of the phase but receive at least one opinion during the phase start supporting an opinion at the end of the phase, chosen u.a.r. (counting multiplicities) from

the received opinions[2]. In other words, each node tries to acquire an opinion during each phase of Stage 1, and, as it eventually receives some opinions, it starts supporting one of them (chosen u.a.r.) from the beginning of the next phase. In particular, decided nodes never change their opinion during the entire stage.

More formally, let ϕ, β, and s be three constants satisfying $\phi > \beta > s$. The rounds of Stage 1 are grouped in $T + 2$ phases with

$$T = \lfloor \log(n/(2s/\varepsilon^2 \log n))/\log(\beta/\varepsilon^2 + 1) \rfloor.$$

Phase 0 takes $s/\varepsilon^2 \log n$ rounds, phase $T + 1$ takes $\phi/\varepsilon^2 \log n$ rounds, and each phase j with $1 \leqslant j \leqslant T$ takes β/ε^2 rounds. We denote with τ_j the end of the last round of phase j.

Let t_u be the first time in which u receives any opinion since the beginning of the protocol (with $t_u = 0$ for the source). Let j_u be the phase of t_u, and let $\mathrm{val}(u)$ be an opinion chosen u.a.r. by u among those that it receives during phase j_u[3]. During the first stage of the protocol each node applies the following rule.

Rule of Stage 1. Each decided node u pushes opinion $\mathrm{val}(u)$ during each round of every phase $j = j_u + 1, ..., T + 1$.

8.2.1.2. *The rule during each phase of Stage 2.* During each phase of Stage 2, every node pushes its opinion at each round of the phase. At the end of the phase, each node that received "enough" opinions takes a random sample[2] of them, and starts supporting the most frequent opinion in that sample (breaking ties u.a.r.).

More formally, the rounds of stage 2 are divided in $T' + 1$ phases with

$$T' = \lceil \log(\sqrt{n/\log n}) \rceil.$$

Each phase j, $0 \leqslant j \leqslant T' - 1$, has length 2ℓ with $\ell = \lceil \alpha_4/\varepsilon^2 \rceil$ for some large-enough constant $\alpha_4 > 0$, and phase T' has length $2\ell'$ with $\ell' = O(\varepsilon^{-2} \log n)$. For any finite multiset A of elements in $\{1, \ldots, k\}$, and any $i \in \{1, \ldots, k\}$, let $\mathrm{occ}(i, A)$ be the number of occurrences of i in A, and let

$$\mathrm{mode}(A) = \{i \in \{1, \ldots, k\} \mid \mathrm{occ}(i, A) \geqslant \mathrm{occ}(j, A) \text{ for every } j \in \{1, \ldots, k\}\}.$$

We then define $\mathrm{maj}(A)$ as the most frequent value in A (breaking ties u.a.r.), i.e., $\mathrm{maj}(A)$ is the r.v. on $\{1, \ldots, k\}$ such that

$$\Pr(\mathrm{maj}(A) = i) = \frac{\mathbb{1}_{\{i \in \mathrm{mode}(A)\}}}{|\mathrm{mode}(A)|}.$$

[2]Note that, in the protocol considered in **[FHK14]**, the choice of each node's new opinion in both stages is based on the first messages received. In **[FHK15]**, in order to relax the synchronicity assumption that nodes share a common clock, they adopt the same sample-based variant of the rule that we adopt here.

[3]Note that, in order to sample u.a.r. one of them, u does not need to collect all the opinions it receives: A natural sampling strategy such as reservoir sampling can be used to this end.

Let $R_j(u)$ be the multiset of messages received by node u during phase j. During the second stage of the protocol each node applies the following rule.

Rule of Stage 2. During each phase j of length $2L$ of Stage 2 ($L = \ell$ or ℓ'), each node u pushes its current opinion at each round of the phase, and starts drawing a random uniform sample $\mathcal{S}(u)$ of size L from $R_j(u)$. Provided $|R_j(u)| \geqslant L$, at the end of the phase u changes its opinion to $\mathrm{maj}(\mathcal{S}(u))$.

Let us remark that the reason we require the use of sampling in the previous rule is that at a given round a node may receive much more messages than $2L$. Thus, if the nodes were to collect all the messages they receive, some of them would need much more memory than the protocol does. Finally, observe that overall both stages 1 and 2 take $O(\frac{1}{\varepsilon^2}\log n)$ rounds.

8.2.2. Pushing Colored Balls into Bins

Before delving into the analysis of the protocol, we provide a framework to rigorously deal with the stochastic dependence that arises between messages in the uniform $\mathcal{PUSH}$ model. Let process $\mathbf{O}$ be the process that results from the execution of the protocol of Section 8.2.1 in the uniform $\mathcal{PUSH}$ model. In order to apply concentration of probability results that requires the involved random variables to be independent, we view the messages as balls, and the nodes as bins, and employ Poisson approximation techniques. More specifically, during each phase j of the protocol, let M_j be the set of messages that are sent to random nodes, and N_j be the set of messages sent *after* the noise has acted on them. (In other words, $N_j = \bigcup_u R_j(u)$). We prove that, at the end of phase j, we can equivalently assume that all the messages M_j have been sent to the nodes according to the following process.

Definition 21 (Associated Balls into Bins Process). The balls-into-bins process $\mathbf{B}$ associated to phase j is the two-step process in which the nodes represent bins and all messages sent in the phase represent colored balls, with each color corresponding to some opinion. Initially, balls are colored according to M_j. At the first step, each ball of color $i \in \{1, \ldots, k\}$ is recolored with color $j \in \{1, \ldots, k\}$ with probability $p_{i,j}$, independently of the other balls. At the second step all balls are thrown to the bins u.a.r. as in a balls-into-bins experiment.

Claim 3. *Given the opinion distribution and the number of active nodes at the beginning of phase j, the probability distribution of the opinion distribution and the number of active nodes at the end of phase j in process $\mathbf{O}$ is the same as if the messages were sent according to process $\mathbf{B}$.*

It is not hard to see that Claim 3 holds in the case of a single round. For more than one round, it is crucial to observe that the way each node u acts in the protocol depends only on the received messages $R_j(u)$, regardless of the order in which these messages are received. As an example, consider the opinion distribution in which one node has opinion 1, one other node

has opinion 2, and all other nodes have opinion 3. Suppose that each node pushes its opinion for two consecutive rounds. Since, at each round, exactly one opinion 1 and exactly one opinion 2 are pushed, no node can receive two 1s during the first round and then two 2s during the second round, i.e. no node can possibly receive the sequence of messages "1,1,2,2" in this exact order. Instead, in process **B** such a sequence is possible.

PROOF OF CLAIM 3. In both process **B** and process **O**, at each round, the noise acts independently on each ball/message of a given color/opinion, according to the same probability distribution for that color/opinion. Then, in both processes, each ball/message is sent to some bin/node chosen u.a.r. and independently of the other balls/messages. Indeed, we can couple process **B** and process **O** by requiring that:

(1) each ball/message is changed by the noise to the same color/value, and
(2) each ball/message ends up in the same bin/node.

Thus, the joint probability distribution of the sets $\{R_j(u)\}_{u\in[n]}$ in process **O** is the same as the one given by process **B**.

Observe also that, from the definition of the protocol (see the rule of Stage 1 and Stage 2 in Section 8.2.1), it follows that each node's action depends only on the set $R_j(u)$ of received messages at the end of each phase j, and does not depend on any further information such as the actual order in which the messages are received during the phase.

Summing up the two previous observations, we get that if, at the end of each phase j, we generate the $R_j(u)$s according to process **B**, and we let the protocol execute according to them, then we indeed get the same stochastic process as process **O**. gg

Now, one key ingredient in our proof is to approximate process **B** using the following process **P**.

DEFINITION 22 (Associated Poisson Process). Given N_j, process **P** associated to phase j is the one-shot process in which each node receives a number of opinions i that is a random variable with distribution Poisson(h_i/n), where h_i is the number of messages in N_j carrying opinion i, and each Poisson random variable is independent of the others.

Now we provide some results from the theory of Poisson approximation for balls-in-bins experiments that are used in Section 8.2.2. For a nice introduction to the topic, we refer to [**MU05**].

LEMMA 54. *Let $\{X_j\}_{j\in[\tilde{n}]}$ be independent r.v. such that $X_j \sim Poisson(\lambda_j)$. The vector $(X_1, ..., X_{\tilde{n}})$ conditional on $\sum_j X = \tilde{m}$ follows a multinomial distribution with $\tilde{m}$ trials and probabilities $(\frac{\lambda_1}{\sum_j \lambda_j}, ..., \frac{\lambda_{\tilde{n}}}{\sum_j \lambda_j})$.*

LEMMA 55. *Consider a balls-in-bins experiment in which h colored balls are thrown in n bins, where h_i balls have color i with $i \in \{1, ..., k\}$ and*

$\sum_i h_i = h$. *Let* $\{X_{u,i}\}_{u\in\{1,...,n\},i\in\{1,...,k\}}$ *be the number of* i*-colored balls that end up in bin* u*, let*

$$f(x_{1,1},...,x_{n,1},x_{n,2},...,x_{n,k},z_1,...,z_n)$$

be a non-negative function with positive integer arguments $x_{1,1},...,x_{n,1},x_{n,2},...,x_{n,k}$, $z_1,...,z_n$*, let* $\{Y_{u,i}\}_{u\in\{1,...,n\},i\in\{1,...,k\}}$ *be independent r.v. such that* $Y_{u,i}\sim$ *Poisson*(h_i/n) *and let* $Z_1,...,Z_n$ *be integer valued r.v. independent from the* $X_{u,i}$*s and* $Y_{u,i}$*s. Then*

$$\mathbb{E}\left[f\left(X_{1,1},...,X_{n,1},X_{n,2},...,X_{n,k},Z_1,...,Z_n\right)\right] \leqslant e^k\sqrt{\prod_i h_i}\,\mathbb{E}\left[f\left(Y_{1,1},...,Y_{n,1},Y_{n,2},...,Y_{n,k},Z_1,...,Z_n\right)\right].$$

Proof. To simplify notation, let

$$\begin{aligned}\bar{Z} &= (Z_1,...,Z_n),\\ \bar{X} &= (X_{1,1},...,X_{n,1},X_{n,2},...,X_{n,n}),\\ \bar{Y} &= (Y_{1,1},...,Y_{n,1},Y_{n,2},...,Y_{n,n}),\\ \bar{Y}_{\sum} &= (\sum_{u=1}^n Y_{u,1},...,\sum_{u=1}^n Y_{u,k}),\\ \lambda_i &= h_i/n,\\ \bar{\lambda} &= (\lambda_1,...,\lambda_k),\end{aligned}$$

and finally $\bar{x}=(x_1,...,x_k)$ for any $x_1,...,x_k$. Observe that, while $X_{u,i}$ and $X_{v,i}$ are clearly dependent, $X_{u,i}$ and $X_{v,j}$ with $i\neq j$ are stochastically independent (even if $u=v$). Indeed, the distribution of the r.v. $\{X_{u,i}\}_{u\in\{1,...,n\}}$ for each fixed i is multinomial with λ_i trials and uniform distribution on the bins. Thus, from Lemma 54 we have that $\{X_{u,i}\}_{u\in\{1,...,n\}}$are distributed as $\{Y_{u,i}\}_{u\in\{1,...,n\}}$ conditional on $\sum_{u=1}^n Y_{u,i}=\lambda_i$, that is

$$\mathbb{E}\left[f\left(\bar{Y},\bar{Z}\right)\middle|\sum_{u=1}^n Y_{u,1}=\lambda_1,...,\sum_{u=1}^n Y_{u,k}=\lambda_k\right]=\mathbb{E}\left[f\left(\bar{X},\bar{Z}\right)\right].$$

Therefore, we have

$$\begin{aligned}\mathbb{E}\left[f\left(\bar{Y},\bar{Z}\right)\right] &= \sum_{\bar{x}:x_1,...,x_k\geqslant 0}\mathbb{E}\left[f\left(\bar{Y},\bar{Z}\right)\middle|\,\bar{Y}_{\sum}=\bar{x}\right]\Pr\left(\bar{Y}_{\sum}=\bar{x}\right)\\ &\geqslant \mathbb{E}\left[f\left(\bar{Y},\bar{Z}\right)\middle|\,\bar{Y}_{\sum}=\bar{\lambda}\right]\Pr\left(\bar{Y}_{\sum}=\bar{\lambda}\right)\\ &= \mathbb{E}\left[f\left(\bar{X},\bar{Z}\right)\right]\Pr\left(\bar{Y}_{\sum}=\bar{\lambda}\right)\\ &= \mathbb{E}\left[f\left(\bar{X},\bar{Z}\right)\right]\prod_i\frac{h_i^{h_i}}{h_i!}e^{-h_i}\geqslant\mathbb{E}\left[f\left(\bar{X},\bar{Z}\right)\right]\frac{e^{-k}}{\sqrt{\prod_i h_i}},\end{aligned}$$

where, in the last inequality, we use that, by Stirling's approximation, $a!\leqslant e\sqrt{a}(\frac{a}{e})^a$ for any $a>0$. ∎

From Lemma 54 and Lemma 55, we get the following general result which says that if a generic event $\mathcal{E}$ holds w.h.p in process $\mathbf{P}$, it also holds in process $\mathbf{O}$, w.h.p.

LEMMA 56. *Given the opinion distribution and the number of active nodes at the beginning of a fixed phase j, let $\mathcal{E}$ be an event that, at the end of that phase, holds with probability at least $1 - n^{-b}$ in process $\mathbf{P}$, for some $b > (k \log h)/(2 \log n)$ with $h = \sum_i h_i$.*[4] *Then, at the end of phase j, $\mathcal{E}$ holds also in process $\mathbf{O}$, w.h.p.*

PROOF. Thanks to Claim 3, it suffices to prove that, at the end of phase j, $\mathcal{E}$ holds in process $\mathbf{B}$, w.h.p.

Let $\bar{\mathcal{E}}$ be the complementary event of $\mathcal{E}$. Let $h = |M_j|$ be the number of balls that are thrown in process $\mathbf{B}$ associated to phase j, where h_i balls have color i with $i \in \{1, ..., k\}$ and $\sum_i h_i = h$. Let $\{X_{u,i}\}_{u \in \{1,...,n\}, i \in \{1,...,k\}}$ be the number of i-colored balls that end up in bin u, let $\{Y_{u,i}\}_{u \in \{1,...,n\}, i \in \{1,...,k\}}$ be the independent r.v. of process $\mathbf{P}$ such that $Y_{u,i} \sim$Poisson(h_i/n) and let $Z_1, ..., Z_n$ be integer valued r.v. independent from the $X_{u,i}$s and $Y_{u,i}$s.

Fix any realization of N_j, i.e. any re-coloring of the balls in the first step of process $\mathbf{B}$. By choosing f in Lemma 55 as the binary r.v. indicating whether event $\bar{\mathcal{E}}$ has occurred, where $\bar{\mathcal{E}}$ is a function of the r.v. $X_{1,1}, ..., X_{n,1}$, $X_{n,2}, ..., X_{n,k}$, $Z_1, ..., Z_n$, we get

$$\begin{aligned} &\Pr\left(\bar{\mathcal{E}}\left(X_{1,1}, ..., X_{n,k}, Z_1, ..., Z_n\right) \middle| N_j\right) \\ &\leqslant e^k \sqrt{\prod_i h_i} \Pr\left(\bar{\mathcal{E}}\left(Y_{1,1}, ..., Y_{n,k}, Z_1, ..., Z_n\right) \middle| N_j\right). \end{aligned} \tag{144}$$

Thus, from (144), the Inequality of arithmetic and geometric means and the hypotheses on the probability of $\mathcal{E}$, we get

$$\begin{aligned} &\Pr\left(\bar{\mathcal{E}}\left(X_{1,1}, ..., X_{n,k}, Z_1, ..., Z_n\right) \middle| N_j\right) \\ &\leqslant e^k \sqrt{\prod_i h_i} \Pr\left(\bar{\mathcal{E}}\left(Y_{1,1}, ..., Y_{n,k}, Z_1, ..., Z_n\right) \middle| N_j\right) \\ &\leqslant e^k \left(\frac{h}{k}\right)^{\frac{k}{2}} \Pr\left(\bar{\mathcal{E}}\left(Y_{1,1}, ..., Y_{n,k}, Z_1, ..., Z_n\right) \middle| N_j\right) \end{aligned}$$

[4] Note that, if N_j is not yet fixed, the parameters h_i of process $\mathbf{P}$ associated to phase j are random variables. However, if the opinion distribution and the number of active nodes at the beginning of phase j are given, then $h = \sum_i h_i = |N_j| = |M_j|$ is fixed.

Finally, let $\mathcal{N}$ be the set of all possible realizations of N_j. By the law of total probability over $\mathcal{N}$, we get that

$$\begin{aligned}
&\sum_{s\in\mathcal{N}} \Pr\left(\bar{\mathcal{E}}\left(X_{1,1},...,X_{n,k},Z_1,...,Z_n\right) \,\middle|\, N_j = s\right) \Pr\left(N_j = s\right) \\
&\leqslant e^k \left(\frac{h}{k}\right)^{\frac{k}{2}} \sum_{s\in\mathcal{N}} \Pr\left(\bar{\mathcal{E}}\left(Y_{1,1},...,Y_{n,k},Z_1,...,Z_n\right) \,\middle|\, N_j = s\right) \Pr\left(N_j = s\right) \\
&\leqslant e^k \left(\frac{h}{k}\right)^{\frac{k}{2}} \Pr\left(\bar{\mathcal{E}}\left(Y_{1,1},...,Y_{n,k},Z_1,...,Z_n\right)\right) \\
&\leqslant \frac{e^k}{k^{\frac{k}{2}}}\, h^{\frac{k}{2}} n^{-b} \leqslant n^{-\Theta(1)},
\end{aligned}$$

where in the first inequality of the last line we used the hypotheses on the probability of $\bar{\mathcal{E}}$. ∎

We now analyze the two stages of our protocol, starting with Stage 1. Note that, in the following two sections, the statements about the evolution of the process refer to process **O**.

8.2.3. Stage 1

The rule of Stage 1 is aimed at guaranteeing that, the system reaches a target opinion distribution from which the bit dissemination problem becomes an instance of the plurality consensus problem, w.h.p. More precisely, we have the following.

LEMMA 57. *Stage 1 takes $O(\frac{1}{\varepsilon^2}\log n)$ rounds, after which all nodes are active and $\mathbf{c}^{(\tau_{T+1})}$ is δ-biased toward the correct opinion with $\delta = \Omega(\sqrt{\log n/n})$, w.h.p.*

PROOF. The fact that an undecided node becomes decided during a phase only depends on whether it gets a message during that phase, regardless of the value of such messages. Hence, the proof that $a^{(\tau_{T+1})} = 1$ is reduced to the analysis of the rule of Stage 1 as an information spreading process, w.h.p. First, by carefully exploiting the Chernoff (Lemma 25) bound and Lemma 56, we can establish Claim 4 and Claim 5 below:

CLAIM 4. *At the end of phase 0, it holds w.h.p.*

$$\frac{s}{\varepsilon^2}\cdot\frac{\log n}{3n} \leqslant a^{(\tau_0)} \leqslant \frac{s}{\varepsilon^2}\cdot\frac{\log n}{n}.$$

CLAIM 5. *At the end of phase j, $1 \leqslant j \leqslant T$, it holds w.h.p.*

$$\left(\frac{\beta}{\varepsilon^2}+1\right)^j \cdot \frac{a^{(\tau_0)}}{8} \leqslant a^{(\tau_j)} \leqslant \left(\frac{\beta}{\varepsilon^2}+1\right)^j \cdot a^{(\tau_0)}.$$

SKETCH OF PROOF OF CLAIM 4 AND CLAIM 5. The probability that, in the process **O**, an undecided node becomes decided at the end of phase j is $1-(1-\frac{1}{n})^h$ where h is the number of messages sent during that phase.

In process $\mathbf{P}$, this probability is $1-e^{-\frac{h}{n}}$. By using that $e^{\frac{x}{1+x}} \leqslant 1+x \leqslant e^x$ for $|x|<1$ we see that

$$1-e^{-\frac{h}{n}} \leqslant 1-\left(1-\frac{1}{n}\right)^h \leqslant 1-e^{-\frac{h}{n-1}}.$$

Thus, we can prove Claim 4 and Claim 5 for process $\mathbf{P}$ by repeating essentially the same calculations as in the proofs of Claim 2.2 and 2.4 in [**FHK15**]. Since the Poisson distributions in process $\mathbf{P}$ are independent, we can apply the Chernoff bound as claimed in [**FHK15**]. Finally, we can prove that the statements hold also for process $\mathbf{O}$ thanks to Lemma 55. gg (of claims 4 and 5)

From the previous two claims, and by the definition of T we get the following.

LEMMA 58. *At the end of phase T, it holds w.h.p.*

$$a^{(\tau_{T+1})} = \Omega\left(\left(\frac{\beta}{\varepsilon^2}+1\right)^T a^{(\tau_0)}\right) = \Omega(\varepsilon^2).$$

Finally, from Lemma 58, an application of the Chernoff bound (Lemma 25) gives us the following.

LEMMA 59. *At the end of Stage 1, all nodes are decided, w.h.p.*

As for the fact that, $\mathbf{c}^{(\tau+1)}$ is a δ-biased opinion distribution with $\delta = \Omega(\sqrt{\log n/n})$ (w.h.p.), we can prove the following.

LEMMA 60. *At the end of each phase j of Stage 1, we have an $(\varepsilon/2)^j$-biased opinion distribution, w.h.p.*

PROOF. We prove the lemma by induction on the phase number. The case $j=1$ is a direct application of Lemma 77 to $c_m^{(\tau_1)} - c_i^{(\tau_1)}$ $(i \neq m)$, where the number of decided nodes is given by Claim 4, and, where the independence of the r.v. follows from the fact that each node that becomes decided in the first phase has necessarily received the messages from the source-node. Now, suppose that the lemma holds up to phase $j-1 \leqslant T$. Let $S_j = \{u|\ j_u = j\}$ be the set of nodes that become decided during phase j. Recall the definition of M_j and N_j from Section 8.2.2, and observe that

$$|M_j| = |N_j| = (\tau_j - \tau_{j-1})\, n \cdot a^{(\tau_{j-1})},$$

and that the number of times opinion i occurs in M_j is $|M_j|\, c_i^{(\tau_{j-1})}$. Let us identify each message in M_j with a distinct number in $1, ..., |M_j|$, and let $\{X_w(i)\}_{w\in\{1,...,|M_j|\}}$ be the binary r.v. such that $X_w(i)=1$ if and only if w is i after the action of the noise. The frequency of opinion i in N_j is $\frac{1}{|N_j|}\sum_{w=1}^{|N_j|} X_w(i)$.

Thanks to Lemma 56, it suffices to prove the lemma for process $\mathbf{P}$. By definition, in process $\mathbf{P}$, for each i, the number of messages with opinion i that each node receives conditional on N_j follows a Poisson$(\frac{1}{n}\sum_{w=1}^{|N_j|} X_w(i))$

distribution. Each node u that becomes decided during phase j gets at least one message during the phase. Thus, from Lemma 54, the probability that u gets opinion i conditional on N_j is

$$\frac{\sum_{w=1}^{|N_j|} X_w(i)}{\sum_{i=1}^{k}\sum_{w=1}^{|N_j|} X_w(i)} = \frac{1}{|N_j|}\sum_{w=1}^{|N_j|} X_w(i).$$

Since decided nodes never change opinion during Stage 1, the bias of $\mathbf{c}^{(\tau_j)}$ is at least the minimum between the bias of $\mathbf{c}^{(\tau_{j-1})}$ and the bias among the newly decided nodes in S_j. Hence, we can apply the Chernoff bound (Lemma 25) to the nodes in S_j to prove that the bias at the end of phase j is w.h.p.[5],

$$\begin{aligned}&\Pr\left(c_m^{(\tau_j)} - c_i^{(\tau_j)}\middle| N_j\right)\\ &\geqslant \left(\frac{1}{|N_j|}\sum_{w=1}^{|N_j|} X_w(m) - \frac{1}{|N_j|}\sum_{w=1}^{|N_j|} X_w(i)\right)\left(1-\tilde{\delta}_j\right),\end{aligned} \tag{145}$$

where $\tilde{\delta}_j = O(\sqrt{\log n/|S_j|})$.

Moreover, note that

$$\mathbb{E}\left[\frac{1}{|N_j|}\sum_{w=1}^{|N_j|} X_w(i)\middle|\, \mathbf{c}^{(\tau_{j-1})}, a^{(\tau_{j-1})}\right] = \left(\mathbf{c}^{(\tau_{j-1})}\cdot P\right)_i.$$

Furthermore, (conditional on $\mathbf{c}^{(\tau_{j-1})}$ and $a^{(\tau_{j-1})}$) the r.v. $\{X_w(i)\}_{w\in\{1,\ldots,|N_j|\}}$ are independent. Thus, for each $i \neq m$, from Claim 5, and by applying the Chernoff bound (Lemma 25) on $\sum_{w=1}^{|N_j|} X_w(m)$, and on $\sum_{w=1}^{|N_j|} X_w(i)$, we get that w.h.p.

$$\frac{1}{|N_j|}\sum_{w=1}^{|N_j|} X_w(m) - \frac{1}{|N_j|}\sum_{w=1}^{|N_j|} X_w(i) \geqslant (1-\delta_j)\,2^{-j+1}\varepsilon^j, \tag{146}$$

where $\delta_j = O(\sqrt{\log n/|N_j|})$.

From Claim 4 and Claim 5, it follows that $\tilde{\delta}_j, \delta_j \leqslant \frac{1}{4}$, w.h.p. Thus by putting together (145) and (146) via the chain rule, we get that w.h.p.

$$c_m^{(\tau_j)} - c_i^{(\tau_j)} \geqslant \left(1-\tilde{\delta}_j\right)(1-\delta_j)\,2^{-j+1}\varepsilon^j \geqslant \left(\frac{\varepsilon}{2}\right)^j.$$

gg (of Lemma 60)

Lemma 60 implies that we get a bias $\varepsilon^{T+2} = \Omega(\sqrt{\log n/n})$ at the end of Stage 1, w.h.p., which completes the proof of Lemma 57. gg

[5]We remark that (145) concerns the value of $\Pr(c_m^{(\tau_j)} - c_i^{(\tau_j)}|N_j)$, which is a random variable.

8.2.4. Stage 2

As proved in the previous section, all nodes are decided at the end of Stage 1 and the final opinion distribution is $\Omega(\sqrt{\log n/n})$-biased, w.h.p. Now, we have that the bit dissemination problem is reduced to an instance of the plurality consensus problem. The purpose of Stage 2 is to progressively amplify the initial bias until all nodes support the plurality opinion, i.e. the opinion originally held by the source node.

During the first T' phases, it is not hard to see that, by taking α_4 large enough, a fraction arbitrarily close to 1 of the nodes receives at least ℓ messages, w.h.p. Each node u in such fraction changes its opinion at the end of the phase. With a slight abuse of notation, let $\mathrm{maj}_\ell(u) = \mathrm{maj}\left(\mathcal{S}(u)\right)$ be u's new opinion based on the $\ell = |\mathcal{S}(u)|$ randomly sampled received messages. We show that these new opinions increase the bias of the opinion distribution toward the plurality opinion by a constant factor > 1, w.h.p.

For the sake of simplicity, we assume that ℓ is odd (see Section 8.5 for details on how to remove this assumption).

PROPOSITION 1. *Suppose that, at the beginning of phase j of Stage 2 with $0 \leqslant j \leqslant T' - 1$, the opinion distribution is δ-biased toward m. In process $\mathbf{P}$, if a node u changes its opinion at the end of the phase, then, for any $i \neq m$, we have*

$$\Pr\left(\mathrm{maj}_\ell(u) = m\right) - \Pr\left(\mathrm{maj}_\ell(u) = i\right) \geqslant \sqrt{\frac{2\ell}{\pi}} \frac{g(\delta,\ell)}{e^{(k-2)\ln 4}}, \tag{147}$$

where

$$g\left(\delta,\ell\right) = \begin{cases} \delta(1-\delta^2)^{\frac{\ell-1}{2}} & \text{if } \delta < \frac{1}{\sqrt{\ell}}, \\ \sqrt{1/\ell}\left(1-\sqrt{1/\ell}\right)^{\frac{\ell-1}{2}} & \text{if } \delta \geqslant \frac{1}{\sqrt{\ell}}. \end{cases}$$

First, we prove (147) for $k = 2$. We then obtain the general case by induction. The proof for $k = 2$ is based on a known relation between the cumulative distribution function of the binomial distribution, and the cumulative distribution function of the beta distribution. This relation is given by the following lemma.

LEMMA 61. *Given $p \in (0,1)$ and $0 \leqslant j \leqslant \ell$ it holds*

$$\sum_{j<i\leqslant\ell} \binom{\ell}{i} p^i \left(1-p\right)^{\ell-i} = \binom{\ell}{j+1} (j+1) \int_0^p z^j \left(1-z\right)^{\ell-j-1} dz.$$

PROOF. By integrating by parts, for $j < \ell - 1$ we have

$$\begin{aligned} &\binom{\ell}{j+1} (j+1) \int_0^p z^j \left(1-z\right)^{\ell-j-1} dz \\ &= \binom{\ell}{j+1} p^{j+1} \left(1-p\right)^{\ell-j-1} \\ &\quad - \binom{\ell}{j+1} (\ell-j-1) \int_0^p z^{j+1} \left(1-z\right)^{\ell-j-2} dz \end{aligned}$$

$$= \binom{\ell}{j+1} p^{j+1} (1-p)^{\ell-j-1}$$

$$- \binom{\ell}{j+2} (j+2) \int_0^p z^{j+1} (1-z)^{\ell-j-2} dz, \tag{148}$$

where, in the last equality, we used the identity

$$\binom{\ell}{j} (\ell - j) = \binom{\ell}{j+1} (j+1).$$

Note that when $j = \ell - 1$, (147) becomes

$$p^\ell = \ell \int_0^p z^{\ell-1} dz.$$

Hence, we can unroll the recurrence given by (148) to obtain

$$\begin{aligned}
&\binom{\ell}{j+1} (j+1) \int_0^p z^j (1-z)^{\ell-j-1} dz \\
&= \sum_{j<i\leqslant \ell-1} \binom{\ell}{i} p^i (1-p)^{\ell-i} + \ell \int_0^p z^{\ell-1} dz \\
&= \sum_{j<i\leqslant \ell} \binom{\ell}{i} p^i (1-p)^{\ell-i},
\end{aligned}$$

concluding the proof. ▣

Lemma 61 allows us to express the survival function of a binomial sample as an integral. Thanks to it, we can prove Proposition 1 when $k = 2$.

LEMMA 62. *Let $\mathbf{c} = (c_1, c_2)$ be a δ-biased opinion distribution during Stage 2. In process $\mathbf{P}$, for any node u, we have*

$$\Pr(\mathrm{maj}_\ell(u) = m) - \Pr(\mathrm{maj}_\ell(u) = 3 - m) \geqslant \sqrt{\frac{2\ell}{\pi}} \cdot g(\delta, \ell).$$

PROOF. Without loss of generality, let $m = 1$. Let $X_1^{(\ell)}$ be a r.v. with distribution $Bin(\ell, p_1)$, and let $X_2^{(\ell)} = \ell - X_1^{(\ell)}$. By using Lemma 61, we get

$$\begin{aligned}
&\Pr(\mathrm{maj}_\ell(u) = 1) - \Pr(\mathrm{maj}_\ell(u) = 2) \\
&= \Pr\left(X_1^{(\ell)} > X_2^{(\ell)}\right) - \Pr\left(X_2^{(\ell)} > X_1^{(\ell)}\right) \\
&= \sum_{\lceil \frac{\ell}{2} \rceil \leqslant i \leqslant \ell} \binom{\ell}{i} p_1^i p_2^{\ell-i} - \sum_{\lceil \frac{\ell}{2} \rceil \leqslant i \leqslant \ell} \binom{\ell}{i} p_1^{\ell-i} p_2^i \\
&= \sum_{\lceil \frac{\ell}{2} \rceil \leqslant i \leqslant \ell} \binom{\ell}{i} p_1^i (1-p_1)^{\ell-i} - \sum_{\lceil \frac{\ell}{2} \rceil \leqslant i \leqslant \ell} \binom{\ell}{i} p_1^{\ell-i} (1-p_1)^i \\
&= \binom{\ell}{\lceil \frac{\ell}{2} \rceil} \left\lceil \frac{\ell}{2} \right\rceil \left(\int_0^{p_1} z^{\lfloor \frac{\ell}{2} \rfloor} (1-z)^{\lfloor \frac{\ell}{2} \rfloor} dz \right.
\end{aligned}$$

$$- \int_0^{p_2} z^{\lfloor \frac{\ell}{2} \rfloor} (1-z)^{\lfloor \frac{\ell}{2} \rfloor} \, dz \Bigg).$$

By setting $t = z - \frac{1}{2}$, and rewriting $p_1 = \frac{p_1 - p_2}{2} + \frac{1}{2}$ and $p_2 = \frac{p_2 - p_1}{2} + \frac{1}{2}$ we obtain

$$\begin{aligned}
&\Pr\left(\mathrm{maj}_\ell(u) = 1\right) - \Pr\left(\mathrm{maj}_\ell(u) = 2\right) \\
&= \binom{\ell}{\lceil \frac{\ell}{2} \rceil} \left\lceil \frac{\ell}{2} \right\rceil \left(\int_0^{p_1} z^{\lfloor \frac{\ell}{2} \rfloor} (1-z)^{\lfloor \frac{\ell}{2} \rfloor} \, dz \right. \\
&\qquad \left. - \int_0^{p_2} z^{\lfloor \frac{\ell}{2} \rfloor} (1-z)^{\lfloor \frac{\ell}{2} \rfloor} \, dz \right) \\
&= \binom{\ell}{\lceil \frac{\ell}{2} \rceil} \left\lceil \frac{\ell}{2} \right\rceil \left(\int_{-\frac{1}{2}}^{\frac{p_1 - p_2}{2}} \left(\frac{1}{4} - t^2 \right)^{\lfloor \frac{\ell}{2} \rfloor} dt \right. \\
&\qquad \left. - \int_{-\frac{1}{2}}^{\frac{-p_1 - p_2}{2}} \left(\frac{1}{4} - t^2 \right)^{\lfloor \frac{\ell}{2} \rfloor} dt \right) \\
&= \binom{\ell}{\lceil \frac{\ell}{2} \rceil} \left\lceil \frac{\ell}{2} \right\rceil \int_{-\frac{p_1 - p_2}{2}}^{\frac{p_1 - p_2}{2}} \left(\frac{1}{4} - t^2 \right)^{\lfloor \frac{\ell}{2} \rfloor} dt.
\end{aligned}$$

For any $t \in (-\frac{y}{2}, \frac{y}{2}) \subseteq (-\frac{p_1 - p_2}{2}, \frac{p_1 - p_2}{2})$, it holds

$$\left(\frac{1}{4} - t^2 \right)^{\lfloor \frac{\ell}{2} \rfloor} \geqslant \left(\frac{1 - y^2}{4} \right)^{\lfloor \frac{\ell}{2} \rfloor}.$$

Thus, for any $y \in (-p_1 + p_2, p_1 - p_2)$ we have

$$\int_{-\frac{p_1 - p_2}{2}}^{\frac{p_1 - p_2}{2}} \left(\frac{1}{4} - t^2 \right)^{\lfloor \frac{\ell}{2} \rfloor} dt \geqslant y \left(\frac{1 - y^2}{4} \right)^{\lfloor \frac{\ell}{2} \rfloor}. \tag{149}$$

The r.h.s. of (149) is maximized w.r.t. $y \in (-p_1 + p_2, p_1 - p_2)$ when

$$y = \min \left\{ p_1 - p_2, \frac{1}{\sqrt{2 \lfloor \frac{\ell}{2} \rfloor + 1}} \right\} = \min \left\{ p_1 - p_2, \frac{1}{\sqrt{\ell}} \right\}.$$

Hence, for $p_1 - p_2 < \frac{1}{\sqrt{\ell}}$, we get

$$\begin{aligned}
&\int_{-\frac{p_1 - p_2}{2}}^{\frac{p_1 - p_2}{2}} \left(\frac{1}{4} - t^2 \right)^{\lfloor \frac{\ell}{2} \rfloor} dt \\
&\geqslant (p_1 - p_2) \left(\frac{1 - (p_1 - p_2)^2}{4} \right)^{\lfloor \frac{\ell}{2} \rfloor} \\
&= 2^{-\ell+1} (p_1 - p_2) \left(1 - (p_1 - p_2)^2 \right)^{\frac{\ell - 1}{2}} \\
&= 2^{-\ell+1} g(p_1 - p_2, \ell).
\end{aligned}$$

For $p_1 - p_2 \geqslant \frac{1}{\sqrt{\ell}}$ we get

$$\int_{-\frac{p_1-p_2}{2}}^{\frac{p_1-p_2}{2}} \left(\frac{1}{4} - t^2\right)^{\left\lfloor \frac{\ell}{2} \right\rfloor} dt \geqslant \frac{2^{-\ell+1}}{\sqrt{\ell}} \left(1 - \frac{1}{\ell}\right)^{\frac{\ell-1}{2}} = 2^{-\ell+1} g\left(p_1 - p_2, \ell\right).$$

By using the fact that g is a non-decreasing function w.r.t. its first argument (see Lemma 68 in Section 8.7), we obtain

$$\begin{aligned}
&\Pr\left(\mathrm{maj}_\ell(u) = 1\right) - \Pr\left(\mathrm{maj}_\ell(u) = 2\right) \\
&= \binom{\ell}{\left\lceil \frac{\ell}{2} \right\rceil} \left\lceil \frac{\ell}{2} \right\rceil \int_{-\frac{p_1-p_2}{2}}^{\frac{p_1-p_2}{2}} \left(\frac{1}{4} - t^2\right)^{\left\lfloor \frac{\ell}{2} \right\rfloor} dt \\
&\geqslant \binom{\ell}{\left\lceil \frac{\ell}{2} \right\rceil} \left\lceil \frac{\ell}{2} \right\rceil 2^{-\ell+1} g\left(p_1 - p_2, \ell\right) \\
&\geqslant \binom{\ell}{\left\lceil \frac{\ell}{2} \right\rceil} \left\lceil \frac{\ell}{2} \right\rceil 2^{-\ell+1} g\left(\delta, \ell\right).
\end{aligned}$$

Finally, by using the bounds $\binom{2r}{r} \geqslant \frac{2^{2r}}{\sqrt{\pi r}} e^{\frac{1}{9r}}$ (see Lemma 69 in Section 8.7), and $e^x \geqslant 1 - x$ together with the identity[6]

$$\binom{\ell}{\left\lceil \frac{\ell}{2} \right\rceil} \left\lceil \frac{\ell}{2} \right\rceil = \binom{\ell}{\frac{\ell+1}{2}} \frac{\ell+1}{2} = \binom{\ell-1}{\frac{\ell-1}{2}} \ell,$$

we get

$$\begin{aligned}
&\Pr\left(\mathrm{maj}_\ell(u) = 1\right) - \Pr\left(\mathrm{maj}_\ell(u) = 2\right) \\
&\geqslant \binom{\ell}{\left\lceil \frac{\ell}{2} \right\rceil} \left\lceil \frac{\ell}{2} \right\rceil 2^{-\ell+1} g\left(\delta, \ell\right) \\
&\geqslant \frac{2^{\ell-1}}{\sqrt{\pi \frac{\ell-1}{2}}} e^{\frac{2}{9(\ell-1)}} \ell \cdot 2^{-\ell+1} g\left(\delta, \ell\right) \\
&\geqslant \sqrt{\frac{2\ell}{\pi}} \left(1 - \frac{2}{9\left(\ell-1\right)}\right) \left(1 - \frac{1}{\ell}\right)^{-\frac{1}{2}} \cdot g\left(\delta, \ell\right) \\
&\geqslant \sqrt{\frac{2\ell}{\pi}} \cdot g\left(\delta, \ell\right),
\end{aligned}$$

concluding the proof. gg

Next we show how to lower bound the above difference with a much simpler expression.

LEMMA 63. *In process* **P**, *during Stage 2, for any node* u,

$$\begin{aligned}
&\Pr(\mathrm{maj}_\ell(u) = m) - \Pr(\mathrm{maj}_\ell(u) = 3 - m) \geqslant \\
&\Pr(X_1^{(\ell)} > X_2^{(\ell)}, ..., X_k^{(\ell)}) - \Pr(X_i^{(\ell)} > X_1^{(\ell)}, ..., X_{i-1}^{(\ell)}, X_{i+1}^{(\ell)}, ..., X_k^{(\ell)}),
\end{aligned}$$

[6]Recall that we are assuming that ℓ is odd.

where $\bar{X}^{(\ell)} = (X_1^{(\ell)}, ..., X_k^{(\ell)})$ *follows a multinomial distribution with* ℓ *trials and probability distribution* $\mathbf{c} \cdot P$.

PROOF. Without loss of generality, let $m = 1$. Let $\mathbf{x} = (x_1, ..., x_k)$ denote a generic vector with positive integer entries such that $\sum_{j=1}^{k} x_j = \ell$, let $W(\mathbf{x})$ be the set of the greatest entries of $\mathbf{x}$, and, for $j \in \{1, i\}$, let

- $A_j^{(!)} = \{\mathbf{x} \,|\, W(\mathbf{x}) = \{j\}\}$,
- $A_j^{(=)} = \{\mathbf{x} \,|\, 1, i \in W(\mathbf{x})\}$,
- $A_1^{(\neq)} = \{\mathbf{x} \,|\, 1 \in W(\mathbf{x}) \wedge i \notin W(\mathbf{x}) \wedge |W(\mathbf{x})| > 1\}$ and
- $A_i^{(\neq)} = \{\mathbf{x} \,|\, i \in W(\mathbf{x}) \wedge 1 \notin W(\mathbf{x}) \wedge |W(\mathbf{x})| > 1\}$.

It holds

$$\begin{aligned}
&\Pr\left(\mathrm{maj}_\ell(u) = j\right) \\
&= \sum_{\mathbf{x} \in A_j^{(!)}} \Pr\left(\bar{X}^{(\ell)} = \mathbf{x}\right) \Pr\left(\mathrm{maj}_\ell(u) = j \,\middle|\, \bar{X}^{(\ell)} = \mathbf{x}\right) \\
&\quad + \sum_{\mathbf{x} \in A_j^{(=)}} \Pr\left(\bar{X}^{(\ell)} = \mathbf{x}\right) \Pr\left(\mathrm{maj}_\ell(u) = j \,\middle|\, \bar{X}^{(\ell)} = \mathbf{x}\right) \\
&\quad + \sum_{\mathbf{x} \in A_j^{(\neq)}} \Pr\left(\bar{X}^{(\ell)} = \mathbf{x}\right) \Pr\left(\mathrm{maj}_\ell(u) = j \,\middle|\, \bar{X}^{(\ell)} = \mathbf{x}\right) \\
&= \sum_{\mathbf{x} \in A_j^{(!)}} \Pr\left(\bar{X}^{(\ell)} = \mathbf{x}\right) + \sum_{\mathbf{x} \in A_j^{(=)}} \frac{\Pr\left(\bar{X}^{(\ell)} = \mathbf{x}\right)}{|W(\mathbf{x})|} \\
&\quad + \sum_{\mathbf{x} \in A_j^{(\neq)}} \frac{\Pr\left(\bar{X}^{(\ell)} = \mathbf{x}\right)}{|W(\mathbf{x})|}. \qquad (150)
\end{aligned}$$

Let

$$\sigma(\mathbf{x}) = (x_i, ..., x_{i-1}, x_1, x_{i+1}, ..., x_k),$$

be the vector function that swaps the entries x_1 and x_i in $\mathbf{x}$. σ is clearly a bijection between the sets $A_1^{(!)}$,$A_1^{(=)}$,$A_1^{(\neq)}$ and $A_i^{(!)}$, $A_i^{(=)}$, $A_i^{(\neq)}$, respectively, namely

$$\sigma : A_1^{(!)} \twoheadrightarrow A_i^{(!)},\ \sigma : A_1^{(=)} \twoheadrightarrow A_i^{(=)},\ \sigma : A_1^{(\neq)} \twoheadrightarrow A_i^{(\neq)},$$

where $\twoheadrightarrow$ denotes a bijection.

Moreover, for all $\mathbf{x} \in A_j^{(=)}$, it holds

$$\Pr\left(\bar{X}^{(\ell)} = \mathbf{x}\right) = \Pr\left(\bar{X}^{(\ell)} = \sigma(\mathbf{x})\right).$$

Therefore

$$\sum_{\mathbf{x}\in A_1^{(=)}} \Pr\left(\bar{X}^{(\ell)}=\mathbf{x}\right) = \sum_{\sigma(\mathbf{x})\in A_1^{(=)}} \Pr\left(\bar{X}^{(\ell)}=\sigma\left(\mathbf{x}\right)\right) = \sum_{\mathbf{x}\in A_i^{(=)}} \Pr\left(\bar{X}^{(\ell)}=\mathbf{x}\right). \tag{151}$$

Furthermore, for all $\mathbf{x}\in A_1^{(\neq)}$, we have

$$\begin{aligned}\Pr\left(\bar{X}^{(\ell)}=\mathbf{x}\right) &= \binom{\ell}{x_1\ \ldots\ x_k} p_1^{x_1}\ldots p_i^{x_i}\ldots p_k^{x_k} \\ &> \binom{\ell}{x_1\ \ldots\ x_k} p_1^{x_i}\ldots p_i^{x_1}\ldots p_k^{x_k} \\ &= \Pr\left(\bar{X}_1^{(\ell)}=\sigma\left(\mathbf{x}\right)\right),\end{aligned} \tag{152}$$

where $\sigma(\mathbf{x})\in A_i^{(\neq)}$. From (152) we thus have that

$$\sum_{\mathbf{x}\in A_1^{(\neq)}} \Pr\left(\bar{X}^{(\ell)}=\mathbf{x}\right) > \sum_{\sigma(\mathbf{x})\in A_1^{(\neq)}} \Pr\left(\bar{X}^{(\ell)}=\sigma\left(\mathbf{x}\right)\right) = \sum_{\mathbf{x}\in A_i^{(\neq)}} \Pr\left(\bar{X}^{(\ell)}=\mathbf{x}\right). \tag{153}$$

From (150), (151) and (153) we finally get

$$\begin{aligned}&\Pr\left(\mathrm{maj}_\ell(u)=1\right)-\Pr\left(\mathrm{maj}_\ell(u)=i\right) \\ &= \sum_{\mathbf{x}\in A_1^{(!)}} \Pr\left(\bar{X}^{(\ell)}=\mathbf{x}\right) + \sum_{\mathbf{x}\in A_1^{(=)}} \frac{\Pr\left(\bar{X}^{(\ell)}=\mathbf{x}\right)}{|W\left(\mathbf{x}\right)|} \\ &+ \sum_{\mathbf{x}\in A_1^{(\neq)}} \frac{\Pr\left(\bar{X}^{(l)}=\mathbf{x}\right)}{|W\left(\mathbf{x}\right)|} - \sum_{\mathbf{x}\in A_i^{(!)}} \Pr\left(\bar{X}^{(\ell)}=\mathbf{x}\right) \\ &- \sum_{\mathbf{x}\in A_i^{(=)}} \frac{\Pr\left(\bar{X}^{(\ell)}=\mathbf{x}\right)}{|W\left(\mathbf{x}\right)|} - \sum_{\mathbf{x}\in A_i^{(\neq)}} \frac{\Pr\left(\bar{X}^{(\ell)}=\mathbf{x}\right)}{|W\left(\mathbf{x}\right)|} \\ &\geqslant \sum_{\mathbf{x}\in A_1^{(!)}} \Pr\left(\bar{X}^{(\ell)}=\mathbf{x}\right) - \sum_{\mathbf{x}\in A_i^{(!)}} \Pr\left(\bar{X}^{(\ell)}=\mathbf{x}\right) \\ &= \Pr\left(W(\bar{X}^{(\ell)})=\{X_1^{(\ell)}\}\right) - \Pr\left(W(\bar{X}^{(\ell)})=\{X_i^{(\ell)}\}\right),\end{aligned}$$

concluding the proof of Lemma 63. gg

Intuitively, Lemma 63 says that the set of events in which a tie occurs among the most frequent opinions in the node's sample of observed messages

does not favor the probability that the node picks the wrong opinion. Thus, by avoiding considering those events, we get a lower bound on

$$\Pr(\mathrm{maj}_\ell(u) = 1) - \Pr(\mathrm{maj}_\ell(u) = i).$$

Thanks to Lemma 63, the proof of (147) reduces to proving the following.

LEMMA 64. *For any fixed k, and with $\bar{X}$ defined as in Lemma 63, we have*

$$\Pr(X_1^{(\ell)} > X_2^{(\ell)}, ..., X_k^{(\ell)}) - \Pr(X_i^{(\ell)} > X_1^{(\ell)}, ..., X_{i-1}^{(\ell)}, X_{i+1}^{(\ell)}, ..., X_k^{(\ell)}) \geqslant \sqrt{2\ell/\pi}\, \frac{g(\delta, \ell)}{4^{k-2}}. \tag{154}$$

PROOF. We prove (154) by induction. Lemma 62 provides us with the base case for $k = 2$. Let us assume that, for $k \leqslant \kappa$, (154) holds. For $k = \kappa+1$, by using the law of total probability, we have

$$\begin{aligned}
&\Pr\left(X_1^{(\ell)} > X_2^{(\ell)}, ..., X_{\kappa+1}^{(\ell)}\right) \\
&\quad - \Pr\left(X_i^{(\ell)} > X_1^{(\ell)}, ..., X_{i-1}^{(\ell)}, X_{i+1}^{(\ell)}, ..., X_{\kappa+1}^{(\ell)}\right) \\
&\geqslant \sum_{h=0}^{\lfloor \frac{\ell}{\kappa+1} \rfloor} \Pr\left(X_1^{(\ell)} > X_2^{(\ell)}, ..., X_{\kappa+1}^{(\ell)} \middle| X_{\kappa+1}^{(\ell)} = h\right) \Pr\left(X_{\kappa+1}^{(\ell)} = h\right) \\
&\quad - \sum_{h=0}^{\lfloor \frac{\ell}{\kappa+1} \rfloor} \Pr\left(X_i^{(\ell)} > X_1^{(\ell)}, ..., X_{i-1}^{(\ell)}, X_{i+1}^{(\ell)}, ..., X_{\kappa+1}^{(\ell)} \middle| X_{\kappa+1}^{(\ell)} = h\right) \\
&\qquad \cdot \Pr\left(X_{\kappa+1}^{(\ell)} = h\right).
\end{aligned} \tag{155}$$

Now, $\arg\max_j\{X_j^{(\ell)}\} = i$ and $X_{\kappa+1}^{(\ell)} \leqslant \left\lfloor \frac{\ell}{\kappa+1} \right\rfloor$ together imply $X_i^{(\ell)} > X_{\kappa+1}^{(\ell)}$. Thus, in the r.h.s. of (155), we have

$$\begin{aligned}
&\Pr\left(X_1^{(\ell)} > X_2^{(\ell)}, ..., X_{\kappa+1}^{(\ell)} \middle| X_{\kappa+1}^{(\ell)} = h\right) \\
&= \Pr\left(X_1^{(\ell)} > X_2^{(\ell)}, ..., X_\kappa^{(\ell)} \middle| X_{\kappa+1}^{(\ell)} = h\right),
\end{aligned}$$

and

$$\begin{aligned}
&\Pr\left(X_i^{(\ell)} > X_1^{(\ell)}, ..., X_{i-1}^{(\ell)}, X_{i+1}^{(\ell)}, ..., X_{\kappa+1}^{(\ell)} \middle| X_{\kappa+1}^{(\ell)} = h\right) \\
&= \Pr\left(X_i^{(\ell)} > X_1^{(\ell)}, ..., X_{i-1}^{(\ell)}, X_{i+1}^{(\ell)}, ..., X_\kappa^{(\ell)} \middle| X_{\kappa+1}^{(\ell)} = h\right).
\end{aligned}$$

Moreover, $X^{(\ell)}$ follows a multinomial distribution with parameters $\mathbf{p}$ and ℓ. Thus $X_k^{(\ell)} = h$ implies that the remaining entries $X_1^{(\ell)}, ..., X_{k-1}^{(\ell)}$ follow a multinomial distribution with $l - h$ trials, and distribution $(\frac{p_1}{1-p_k}, ..., \frac{p_{k-1}}{1-p_k})$.

Let $Y^{(\ell-h)} = (Y_1^{(\ell-h)}, ..., Y_{k-1}^{(\ell-h)})$ be the distribution of $X_1^{(\ell)}, ..., X_{k-1}^{(\ell)}$ conditional on $X_k^{(\ell)} = h$. From (155) we get

$$\begin{aligned}
&\Pr\left(X_1^{(\ell)} > X_2^{(\ell)}, ..., X_{\kappa+1}^{(\ell)}\right) \\
&\quad - \Pr\left(X_i^{(\ell)} > X_1^{(\ell)}, ..., X_{i-1}^{(\ell)}, X_{i+1}^{(\ell)}, ..., X_{\kappa+1}^{(\ell)}\right) \\
&\geqslant \sum_{h=0}^{\lfloor \frac{\ell}{\kappa+1} \rfloor} \Pr\left(X_1^{(\ell)} > X_2^{(\ell)}, ..., X_\kappa^{(l)} \middle| X_{\kappa+1}^{(\ell)} = h\right) \Pr\left(X_{\kappa+1}^{(\ell)} = h\right) \\
&\quad - \sum_{h=0}^{\lfloor \frac{\ell}{\kappa+1} \rfloor} \Pr\left(X_i^{(\ell)} > X_1^{(\ell)}, ..., X_{i-1}^{(\ell)}, X_{i+1}^{(\ell)}, ..., X_\kappa^{(\ell)} \middle| X_{\kappa+1}^{(\ell)} = h\right) \\
&\qquad \cdot \Pr\left(X_{\kappa+1}^{(\ell)} = h\right) \\
&\geqslant \sum_{h=0}^{\lfloor \frac{\ell}{\kappa+1} \rfloor} \left(\Pr\left(Y_1^{(\ell-h)} > Y_2^{(\ell-h)}, ..., Y_\kappa^{(\ell-h)}\right) - \right. \\
&\quad \left. - \Pr\left(Y_i^{(l-h)} > Y_1^{(l-h)}, ..., Y_{i-1}^{(l-h)}, Y_{i+1}^{(l-h)}, ..., Y_\kappa^{(l-h)}\right)\right) \\
&\qquad \cdot \Pr\left(X_{\kappa+1}^{(l)} = h\right). \qquad (156)
\end{aligned}$$

Now, using the inductive hypothesis on the r.h.s. of (156) we get

$$\begin{aligned}
&\sum_{h=0}^{\lfloor \frac{\ell}{\kappa+1} \rfloor} \left(\Pr\left(Y_1^{(\ell-h)} > Y_2^{(\ell-h)}, ..., Y_\kappa^{(\ell-h)}\right)\right. \\
&\quad \left. - \Pr\left(Y_i^{(\ell-h)} > Y_1^{(\ell-h)}, ..., Y_{i-1}^{(\ell-h)}, Y_{i+1}^{(\ell-h)}, ..., Y_\kappa^{(\ell-h)}\right)\right) \\
&\qquad \cdot \Pr\left(X_{\kappa+1}^{(\ell)} = h\right) \\
&\geqslant \sum_{h=0}^{\lfloor \frac{\ell}{\kappa+1} \rfloor} \left(\sqrt{\frac{2\ell - 2h}{\pi}} \frac{g(\delta, \ell - h)}{4^{\kappa-2}}\right) \Pr\left(X_{\kappa+1}^{(\ell)} = h\right) \\
&\geqslant \sqrt{\frac{2\ell}{\pi}} \frac{g(\delta, \ell)}{4^{\kappa-2}} \cdot \sum_{h=0}^{\lfloor \frac{\ell}{\kappa+1} \rfloor} \sqrt{1 - \frac{h}{\ell}} \Pr\left(X_{\kappa+1}^{(\ell)} = h\right),
\end{aligned}$$

where, in the last inequality, we used the fact that g is a non-increasing function w.r.t. the second argument (see Lemma 68 in Section 8.7).

It remains to show that

$$\sum_{h=0}^{\lfloor \frac{\ell}{\kappa+1} \rfloor} \sqrt{1 - \frac{h}{l}} \Pr\left(X_{\kappa+1}^{(\ell)} = h\right) \geqslant \frac{1}{4}.$$

Let $W_{\kappa+1}^{(\ell)}$ be a r.v. with probability distribution $Bin(\ell, \frac{1}{\kappa+1})$. Since $X_{\kappa+1}^{(\ell)} \sim Bin(\ell, p_{\kappa+1})$ with $p_{\kappa+1} \leqslant \frac{1}{\kappa+1}$, a standard coupling argument (see for example [**DP09**, Exercise 1.1.]), enables to show that

$$\Pr\left(X_{\kappa+1}^{(\ell)} \leqslant h\right) \geqslant \Pr\left(W_{\kappa+1}^{(\ell)} \leqslant h\right).$$

Hence, we can apply the central limit theorem[7] on $W_{\kappa+1}^{(\ell)}$, and get that, for any $\tilde{\varepsilon} \leqslant \frac{2-\sqrt{3}}{4}$, there exists some fixed constant ℓ_0 such that, for $\ell \geqslant \ell_0$, we have

$$\Pr\left(X_{\kappa+1}^{(\ell)} \leqslant \frac{\ell}{\kappa+1}\right) \geqslant \Pr\left(W_{\kappa+1}^{(\ell)} \leqslant \frac{\ell}{\kappa+1}\right) \geqslant \left(\frac{1}{2} - \tilde{\varepsilon}\right). \tag{157}$$

By using (157), for $\ell \geqslant \ell_0$ we finally get that

$$\begin{aligned}
&\sum_{h=0}^{\lfloor \frac{\ell}{\kappa+1} \rfloor} \sqrt{1 - \frac{h}{\ell}} \Pr\left(X_{\kappa+1}^{(\ell)} = h\right) \\
&\geqslant \sqrt{1 - \frac{\left\lfloor \frac{\ell}{\kappa+1} \right\rfloor}{\ell}} \cdot \sum_{h=0}^{\lfloor \frac{\ell}{\kappa+1} \rfloor} \Pr\left(X_{\kappa+1}^{(\ell)} = h\right) \\
&\geqslant \sqrt{1 - \frac{2}{\kappa+1}} \cdot \Pr\left(X_{\kappa+1}^{(\ell)} \leqslant \frac{\ell}{\kappa+1}\right) \\
&\geqslant \sqrt{\frac{\kappa-1}{\kappa+1}} \cdot \left(\frac{1}{2} - \tilde{\varepsilon}\right) \geqslant \sqrt{\frac{1}{3}} \cdot \left(\frac{1}{2} - \tilde{\varepsilon}\right) \geqslant \frac{1}{4},
\end{aligned}$$

concluding the proof that

$$\Pr\left(\mathrm{maj}_\ell(u) = 1\right) - \Pr\left(\mathrm{maj}_\ell(u) = i\right) \geqslant \sqrt{\frac{2\ell}{\pi}} \frac{g\left(\delta, \ell\right)}{e^{(k-2)\ln 4}}.$$

gg

By using Proposition 1, we can then prove Lemma 65.

LEMMA 65. *At the end of Stage 2, all nodes support the initial plurality opinion, w.h.p.*

PROOF. Let $\delta = \Omega(\sqrt{\log n / n})$ be the bias of the opinion distribution at the beginning of a generic phase $j < T'$ of Stage 2. Thanks to Proposition

[7] Recall that the central limit theorem states that given a random sample $X_1, ..., X_n$ from a Bernoulli(p) distribution where $p \in (0,1)$ is constant (i.e. does not depend on n), and given a standard normal r.v. $Z \sim N(0,1)$, it holds

$$\lim_{n\to\infty} \sup_{z\in\mathbb{R}} \left| \Pr\left(\frac{\sum_{i=1}^n X_i - pn}{\sqrt{n}} \leqslant z \right) - \Pr\left(Z \leqslant \frac{z}{\sqrt{p\left(1-p\right)}} \right) \right| = 0.$$

1, by choosing the constant α_4 of the phase length large enough, in process $\mathbf{P}$ we get that

$$\Pr\left(\mathrm{maj}_\ell(u)=m\right)-\Pr\left(\mathrm{maj}_\ell(u)=i\right)\geqslant\alpha\delta$$

for some constant $\alpha>1$ (provided that $\delta\leqslant 1/2$). Hence, by applying Lemma 77 in Appendix A with $\theta=\frac{\alpha}{4}\delta$, we get

$$\Pr(c_m^{(\tau_j)}-c_i^{(\tau_j)}\leqslant\alpha\delta/2)\leqslant\exp(-(\alpha\delta)^2 n/16)\leqslant n^{-\tilde{\alpha}},$$

for some constant $\tilde{\alpha}$ that is large enough to apply Lemma 55. Therefore, until $\delta\geqslant 1/2$, in process $\mathbf{P}$ we have that $c_m^{(\tau_j)}-c_i^{(\tau_j)}\geqslant\alpha\delta/2$ holds, w.h.p. From the previous equation it follows that, after T' phases, the protocol has reached an opinion distribution with a bias greater than $1/2$. Thus, by a direct application of Lemma 77 and Lemma 55 to $c_m^{(\tau_{T'})}-c_i^{(\tau_{T'})}$, we get that w.h.p. $c_m^{(\tau_{T'})}-c_i^{(\tau_{T'})}=1$, concluding the proof. ∎

Finally, the time efficiency claimed in Theorem 13 and Theorem 14 directly follows from Lemma 65, while the required memory follows from the fact that in each phase each node needs only to count how many times it has received each opinion, i.e. to count up to at most $O(\frac{1}{\varepsilon^2}\log n)$, w.h.p.

8.3. On the Notion of (ε,δ)-Majority-Preserving Matrix

In this section we discuss the notion of (ε,δ)-m.p. noise matrix given in Definition 20. Let us consider (143). The matrix P represents the "perturbation" introduced by the noise, and so $(\mathbf{c}\cdot P)_m-(\mathbf{c}\cdot P)_i$ measures how much information the system is losing about the correct opinion m, in a single communication round. An (ε,δ)-m.p. noise matrix is a noise matrix that preserves at least an ε fraction of bias, provided the initial bias is at least δ. The (ε,δ)-m.p. property essentially characterizes the amount of noise beyond which some coordination problems cannot be solved without further hypotheses on the nodes' knowledge of the matrix P. To see why this is the case, consider an (ε,δ)-m.p. noise matrix for which there is a δ-biased opinion distribution $\tilde{\mathbf{c}}$ such that $(\tilde{\mathbf{c}}\cdot P)_m-(\tilde{\mathbf{c}}\cdot P)_i<0$ for some opinion i. Given opinion distribution $\tilde{\mathbf{c}}$, *from each node's perspective, opinion m does not appear to be the most frequent opinion.* Indeed, the messages that are received are more likely to be i than m. Thus, plurality consensus cannot be solved from opinion distribution $\tilde{\mathbf{c}}$.

Observe that verifying whether a given matrix P is (ε,δ)-m.p. with respect to opinion m consists in checking whether for each $i\neq m$ the value of the following linear program is at least $\varepsilon\delta$:

$$\begin{aligned}&\text{maximize } (P\cdot\mathbf{c})_m-(P\cdot\mathbf{c})_i\\&\text{subject to } \sum_j c_j=1,\\&\qquad\text{and } \forall j,\ c_j\geqslant 0, c_m-c_j-\delta\geqslant 0.\end{aligned}$$

We now provide some negative and positive examples of (ε, δ)-m.p. noise matrices. First, we note that a natural matrix property such as being diagonally dominant does not imply that the matrix is (ε, δ)-m.p. For example, by multiplying the following diagonally dominant matrix by the δ-biased opinion distribution $\mathbf{c} = (1/2 + \delta, 1/2 - \delta, 0)^\intercal$, we see that it does not even preserve the majority opinion at all when $\varepsilon, \delta < 1/6$:

$$\begin{pmatrix} \frac{1}{2}+\varepsilon & 0 & \frac{1}{2}-\varepsilon \\ \frac{1}{2}-\varepsilon & \frac{1}{2}+\varepsilon & 0 \\ 0 & \frac{1}{2}-\varepsilon & \frac{1}{2}+\varepsilon \end{pmatrix}.$$

On the other hand, the following natural generalization of the noise matrix in [**FHK14**] (see (142)), is (ε, δ)-m.p. for every $\delta > 0$ with respect to any opinion:

$$(P)_{i,j} = p_{i,j} = \begin{cases} \frac{1}{k} + \varepsilon & \text{if } i = j, \\ \frac{1}{k} - \frac{\varepsilon}{k-1} & \text{otherwise.} \end{cases}$$

More generally, let P be a noise matrix such that

$$(P)_{i,j} = \begin{cases} p & \text{if } i = j, \\ q_l \leqslant q_{i,j} \leqslant q_u & \text{otherwise,} \end{cases} \tag{158}$$

for some positive numbers p, q_u and q_l. Since

$$\begin{aligned} (P\mathbf{c})_m - (P\mathbf{c})_i &= pc_m + \sum_{j \neq m} q_{j,m} c_j - pc_i - \sum_{j \neq i} q_{j,i} c_j \\ &\geqslant p(c_m - c_i) + \sum_{j \neq m} q_l c_j - \sum_{j \neq i} q_u c_j \\ &\geqslant p(c_m - c_i) + q_l(1 - c_m) - q_u(1 - c_i) \\ &\geqslant p(c_m - c_i) + q_l - q_l c_m - q_u + q_u c_i \\ &\geqslant p(c_m - c_i) - q_u(c_m - c_i) - (q_u - q_l) \\ &\geqslant (p - q_u)(c_m - c_i) - (q_u - q_l) \\ &\geqslant (p - q_u)\delta - (q_u - q_l). \end{aligned} \tag{159}$$

By defining $\varepsilon = (p - q_u)/2$, we get that the last line in (159) is greater than $\varepsilon\delta$ iff $(p - q_u)\delta/2 \geqslant (q_u - q_l)$, which gives a sufficient condition for any matrix of the form given in (158) for being (ε, δ)-m.p.

8.4. The Reception of Simultaneous Messages

In the uniform $\mathcal{PUSH}$ model, it may happen that several agents push a message to the same node u at the same round. In such cases, the model should specify whether the node receives all such messages, only one of them or neither of them. Which choice is better depends on the biological setting that is being modeled: if the communication between the agents of the system is an auditory or tactile signal, it could be more realistic to assume that simultaneous messages to the same node would

"collide", and the node would not be able to grasp any of them. If, on the other hand, the messages represent visual or chemical signals (see e.g. [**SKJ+08, FPM+02, BSDDS10, BCN+14**]), then it may be unrealistic to assume that nodes cannot receive more than one of such messages at the same round and besides, by a standard balls-into-bins argument (e.g. by applying Lemma 56), it follows that in the uniform $\mathcal{PUSH}$ model at each round no node receives more than $\mathcal{O}(\log n)$ messages, w.h.p. In this work we thus consider the model in which all messages are received, also because such assumption allows us to obtain simpler proofs than the other variants. We finally note that our protocol does not strictly need such assumption, since it only requires the nodes to collect a small random sample of the received messages. However, since we look at the latter feature as a consequence of active choices of the nodes rather than some inherent property of the environment, we avoid to weaken the model to the point that it matches the requirements of the protocol.

8.5. Removing the Parity Assumption on ℓ

The next lemma shows that, for $k = 2$, the increment of bias at the end of each phase of Stage 2 in the process $\mathbf{P}$ is non-decreasing in the value of ℓ, regardless of its parity.

LEMMA 66. *Let $k = 2$, $a = 1$, let ℓ be odd, and let*

$$(\mathbf{c} \cdot P)_1 \geqslant 1 - (\mathbf{c} \cdot P)_1 = (\mathbf{c} \cdot P)_2.$$

The rule of Stage 2 of the protocol is such that

$$\begin{aligned} \Pr\left(\mathrm{maj}_\ell(u) = 1\right) &= \Pr\left(\mathrm{maj}_{\ell+1}(u) = 1\right) \\ &\leqslant \Pr\left(\mathrm{maj}_{\ell+2}(u) = 1\right), \\ \Pr\left(\mathrm{maj}_\ell(u) = 2\right) &= \Pr\left(\mathrm{maj}_{\ell+1}(u) = 2\right) \\ &\geqslant \Pr\left(\mathrm{maj}_{\ell+2}(u) = 2\right). \end{aligned}$$

Since we are not using any feature of the protocol other than the majority rule, we obtain Lemma 66 as corollary of a general Lemma 67, which is of independent interest. To get Lemma 66 from Lemma 67, set "$\mathrm{maj}_\ell = H$" = "$\mathrm{maj}_\ell(u) = 1$", "$\mathrm{maj}_\ell = T$" = "$\mathrm{maj}_\ell(u) = 2$" and $p = (\mathbf{c} \cdot P)_1$.

LEMMA 67. *Suppose with throw ℓ times a coin whose probability of head is $p \geqslant 1 - p$. Let maj_ℓ be the face of the coin which shows up more frequently in the ℓ throws (i.e. the majority value), breaking ties uniformly at random: if we get $\frac{\ell}{2}$ heads and $\frac{\ell}{2}$ tails, we choose one of them with probability $\frac{1}{2}$ (notice*

that a tie is only possible if ℓ is even. For any odd ℓ, it holds

$$\Pr\left(\mathrm{maj}_\ell = H\right) = \Pr\left(\mathrm{maj}_{\ell+1} = H\right) \tag{160}$$
$$\overset{(a)}{\leqslant} \Pr\left(\mathrm{maj}_{\ell+2} = H\right),$$
$$\Pr\left(\mathrm{maj}_\ell = T\right) = \Pr\left(\mathrm{maj}_{\ell+1} = T\right) \tag{161}$$
$$\overset{(b)}{\geqslant} \Pr\left(\mathrm{maj}_{\ell+2} = T\right),$$

where the equality in (a) and (b) holds iff $p = \frac{1}{2}$.

Proof. By definition, we have

$$\Pr\left(\mathrm{maj}_\ell = H\right) = \Pr\left(X_H^{(\ell)} \geqslant \left\lceil \frac{\ell}{2} \right\rceil\right),$$
$$\Pr\left(\mathrm{maj}_{\ell+1} = H\right) = \Pr\left(X_H^{(\ell+1)} > \frac{\ell+1}{2}\right) + \frac{1}{2}\Pr\left(X_H^{(\ell+1)} = \frac{\ell+1}{2}\right),$$
$$\Pr\left(\mathrm{maj}_{\ell+2} = H\right) = \Pr\left(X_H^{(\ell+2)} \geqslant \left\lceil \frac{\ell+2}{2} \right\rceil\right),$$

where $X_1^{(\ell)}$, $X_1^{(\ell+1)}$ and $X_1^{(\ell+2)}$ are binomial r.v. with probability p_1 and number of trials ℓ, $\ell+1$, and $\ell+2$, respectively. We can view $X_1^{(\ell)}$, $X_1^{(\ell+1)}$, and $X_1^{(\ell+2)}$ as the sum of ℓ, $\ell+1$ and $\ell+2$ Bernoulli(p) r.v., respectively. In particular, let Y and Y' be independent r.v. with distribution Bernoulli(p). We can couple $X_1^{(\ell)}$, $X_1^{(\ell+1)}$ and $X_1^{(\ell+2)}$ as follows:

$$X_1^{(\ell+1)} = X_1^{(\ell)} + Y,$$

and

$$X_1^{(\ell+2)} = X_1^{(\ell+1)} + Y'.$$

Since ℓ is odd, observe that if $X_H^{(\ell)} > \lceil \frac{\ell}{2} \rceil$, then $\mathrm{maj}_\ell = H$ regardless of the value of Y, and similarly if $X_H^{(\ell)} < \lceil \frac{\ell}{2} \rceil$ then $\mathrm{maj}_\ell = T$. Thus we have

$$\begin{aligned}
&\Pr\left(\mathrm{maj}_{\ell+1} = H\right) \\
&= \sum_{i=1}^{\ell} \Pr\left(\mathrm{maj}_{\ell+1} = H \middle| X_H^{(\ell)} = i\right) \Pr\left(X_H^{(\ell)} = i\right) \\
&= \sum_{i > \lceil \frac{\ell}{2} \rceil}^{\ell} \Pr\left(X_H^{(\ell)} = i\right) + \Pr\left(\mathrm{maj}_{\ell+1} = H \middle| X_H^{(\ell)} = \left\lceil \frac{\ell}{2} \right\rceil\right) \\
&\qquad \cdot \Pr\left(X_H^{(\ell)} = \left\lceil \frac{\ell}{2} \right\rceil\right)
\end{aligned}$$

$$+ \Pr\left(\mathrm{maj}_{\ell+1} = H \middle| X_H^{(\ell)} = \left\lfloor \frac{\ell}{2} \right\rfloor\right) \Pr\left(X_H^{(\ell)} = \left\lfloor \frac{\ell}{2} \right\rfloor\right). \tag{162}$$

As for the last two terms in the previous equation, we have that

$$\begin{aligned} &\Pr\left(\mathrm{maj}_{\ell+1} = H \middle| X_H^{(\ell)} = \left\lceil \frac{\ell}{2} \right\rceil\right) \\ &= \Pr(Y = H) + \Pr(Y = T) \cdot \frac{1}{2}, \end{aligned} \tag{163}$$

and

$$\Pr\left(\mathrm{maj}_{\ell+1} = H \middle| X_H^{(\ell)} = \left\lfloor \frac{l}{2} \right\rfloor\right) = \Pr(Y = H) \cdot \frac{1}{2}. \tag{164}$$

Moreover, by a direct calculation one can verify that

$$\Pr\left(X_H^{(\ell)} = \left\lceil \frac{\ell}{2} \right\rceil\right) = \frac{\Pr(Y = T)}{\Pr(Y = H)} \cdot \Pr\left(X_H^{(\ell)} = \left\lfloor \frac{\ell}{2} \right\rfloor\right). \tag{165}$$

From (163), (164) and (165) it follows that

$$\begin{aligned} &\Pr\left(\mathrm{maj}_{\ell+1} = H \middle| X_H^{(\ell)} = \left\lceil \frac{\ell}{2} \right\rceil\right) \Pr\left(X_H^{(\ell)} = \left\lceil \frac{\ell}{2} \right\rceil\right) \\ &\quad + \Pr\left(\mathrm{maj}_{\ell+1} = H \middle| X_H^{(\ell)} = \left\lfloor \frac{\ell}{2} \right\rfloor\right) \Pr\left(X_H^{(\ell)} = \left\lfloor \frac{\ell}{2} \right\rfloor\right) \\ &= \left(\Pr(Y = H) + \Pr(Y = T) \cdot \frac{1}{2}\right) \Pr\left(X_H^{(\ell)} = \left\lceil \frac{\ell}{2} \right\rceil\right) \\ &\qquad + \left(\Pr(Y = H) \cdot \frac{1}{2}\right) \Pr\left(X_H^{(\ell)} = \left\lfloor \frac{\ell}{2} \right\rfloor\right) \\ &= \Pr\left(X_H^{(\ell)} = \left\lceil \frac{\ell}{2} \right\rceil\right) \\ &\qquad \cdot \left(\Pr(Y = H) + \frac{\Pr(Y = T)}{2} + \frac{\Pr(Y = H)}{2} \frac{\Pr(Y = T)}{\Pr(Y = H)}\right) \\ &= \Pr\left(X_H^{(\ell)} = \left\lceil \frac{\ell}{2} \right\rceil\right). \end{aligned} \tag{166}$$

By plugging (166) in (162) we get

$$\Pr(\mathrm{maj}_\ell = H) = \Pr\left(\mathrm{maj}_{\ell+1} = H\right).$$

As for the second part, observe that if $X_H^{(\ell+1)} > \frac{\ell+1}{2}$, then $\mathrm{maj}_{\ell+2} = H$ regardless of the value of Y', and similarly if $X_H^{(\ell+1)} < \frac{\ell+1}{2}$ then $\mathrm{maj}_{\ell+2} = T$. Observe also that

$$\Pr\left(\mathrm{maj}_{\ell+2} = H \middle| X_H^{(\ell+1)} = \frac{l+1}{2}\right) = \Pr(Y = H) = p.$$

Because of the previous observations and the hypothesis that $p \geqslant \frac{1}{2}$, we have that

$$\begin{aligned}
&\Pr\left(\mathrm{maj}_{\ell+2} = H\right) \\
&= \sum_{i=0}^{\ell} \Pr\left(\mathrm{maj}_{\ell+2} = H \middle| X_H^{(\ell+1)} = i\right) \Pr\left(X_H^{(\ell+1)} = i\right) \\
&= \sum_{i > \frac{\ell+1}{2}}^{l} \Pr\left(X_H^{(\ell+1)} = i\right) \\
&\qquad + \Pr\left(\mathrm{maj}_{\ell+2} = H \middle| X_H^{(\ell+1)} = \frac{\ell+1}{2}\right) \\
&\qquad\qquad \cdot \Pr\left(X_H^{(\ell+1)} = \frac{\ell+1}{2}\right) \\
&= \sum_{i > \frac{\ell+1}{2}}^{\ell} \Pr\left(X_H^{(\ell+1)} = i\right) + p \cdot \Pr\left(X_H^{(\ell+1)} = \frac{\ell+1}{2}\right) \\
&\overset{(a)}{\geqslant} \sum_{i > \frac{\ell+1}{2}}^{\ell} \Pr\left(X_H^{(\ell+1)} = i\right) + \frac{1}{2} \cdot \Pr\left(X_H^{(\ell+1)} = \frac{\ell+1}{2}\right) \\
&= \Pr\left(\mathrm{maj}_{\ell+1} = H\right),
\end{aligned}$$

where equality in (a) holds iff $p = \frac{1}{2}$.

Finally, (161) follows from (160) and the fact that

$$\Pr\left(\mathrm{maj}_{\ell} = T\right) = 1 - \Pr\left(\mathrm{maj}_{\ell} = H\right).$$

gg

8.6. Bit dissemination with $\varepsilon = \Theta(n^{-\frac{1}{4}-\eta})$

In [**FHK15**] it is shown that at the end of Stage 1 the bias toward the correct opinion is at least $\varepsilon^{T+2}/2$ and, at the beginning of Stage 2, they assume a bias toward the correct opinion of $\Omega(\sqrt{\log n/n})$. In this section, we show that, when $\varepsilon = \Theta(n^{-\frac{1}{4}-\eta})$ for some $\eta \in (0, 1/4)$, the protocol considered by [**FHK15**] and us cannot solve the bit dissemination and the plurality consensus problem in time $\Theta(\log n/\varepsilon^2)$.

First, observe that when $\varepsilon = \Theta(\sqrt{\log n/n})$ the length of the first phase of Stage 1 is $\Theta\left(\log n/\varepsilon^2\right) = \Omega(n \log n)$, which implies that each node gets at least one message from the source during the first phase, w.h.p. Thus, thanks to our analysis of Stage 2 we have that when $\varepsilon = \Theta(\sqrt{\log n/n})$ the protocol effectively solves the bit dissemination problem in time $\Theta(\log n/\varepsilon^2)$, w.h.p.

In general, for $\varepsilon < n^{-1/2-\eta}$ for some constant $\eta > 0$, if we adopt the second stage right from the beginning (which means that the source node sends

ε^{-2} messages), we get that all nodes receive at least $\log n/(\varepsilon^2 n)$ messages, w.h.p. Thus, by a direct application of Lemma 77, after the first phase we get an $\sqrt{\log n/n}$-biased opinion distribution and Stage 2 correctly solves the problem according to Theorem 14, w.h.p.

However, when $\varepsilon = \Theta(n^{-\frac{1}{4}-\eta})$ for some $\eta > 0$, from Claim 4 and Lemma 60 we have that, after phase 0 in opinion distribution $\mathbf{c}$, at most $\mathcal{O}\left(\log n/\varepsilon^2\right) = \mathcal{O}(n^{\frac{1}{2}+2\eta}\log n)$ nodes are decided, and $\mathbf{c}$ is $\frac{\varepsilon}{2}$-biased. Each node that becomes decided in phase 1 receives a message pushed from some node of $\mathbf{c}$, and, because of the noise, the value of this message is distributed according to $\mathbf{c}^{(\tau_0)} \cdot P$. It follows that $\mathbf{c}$ is an $\varepsilon^2/2$-biased opinion distribution with $\varepsilon^2 = n^{-\frac{1}{2}-2\eta}$ which is much smaller than the $\Omega(\sqrt{\log n/n})$ bound required for the second stage.

We believe that no minor modification of the protocol proposed here can correctly solve the noisy bit dissemination problem when $\varepsilon = \Theta(n^{-\frac{1}{4}-\eta})$ in time $\mathcal{O}\left(\log n/\varepsilon^2\right)$.

8.7. Technical Lemmas

Next lemmas establishes the monotonicity of the function g and an approximation of the binomial coefficient, which have been used in the proof of Lemma 64.

LEMMA 68. *The function*

$$g(x,y) = \begin{cases} x\left(1-x^2\right)^{\frac{y-1}{2}} & \text{if } x < \frac{1}{\sqrt{y}}, \\ \frac{1}{\sqrt{y}}\left(1-\frac{1}{y}\right)^{\frac{y-1}{2}} & \text{if } x \geqslant \frac{1}{\sqrt{y}}, \end{cases}$$

with $x \in [0,1]$ and $y \in [1,+\infty)$ is non-decreasing w.r.t. x and non-increasing w.r.t. y.

PROOF. To show that $g(x,y)$ is non-decreasing w.r.t. x, observe that

$$\frac{\partial}{\partial x} g(x,y) = \left(\left(1-x^2\right)^{\frac{y-1}{2}} - 2x^2\left(\frac{y-1}{2}\right)\left(1-x^2\right)^{\frac{y-1}{2}}\right)$$

for $x < y^{-\frac{1}{2}} < 1$, and

$$\left(1-x^2\right)^{\frac{y-1}{2}} - 2x^2\left(\frac{y-1}{2}\right)\left(1-x^2\right)^{\frac{y-1}{2}} \geqslant 0$$

for $x < y^{-\frac{1}{2}}$.

To show that $g(x,y)$ is non-increasing w.r.t. y, observe that this is true for $x < y^{-\frac{1}{2}}$. For $x \geqslant y^{-\frac{1}{2}}$, since

$$\begin{aligned} &\frac{\partial}{\partial y}\left(\log y^{-\frac{1}{2}} + \frac{y-1}{2}\log\left(1-\frac{1}{y}\right)\right) \\ &= \frac{\partial}{\partial y}\left(\frac{y-1}{2}\log(y-1) - \frac{y}{2}\log y\right) \leqslant 0, \end{aligned}$$

we have

$$\frac{\partial}{\partial y} g\left(x, y\right) = \frac{\partial}{\partial y} \exp\left\{\log y^{-\frac{1}{2}} + \frac{y-1}{2} \log\left(1 - \frac{1}{y}\right)\right\} \leqslant 0,$$

concluding the proof. ∎

LEMMA 69. *For any integer $r \geqslant 1$ it holds*

$$\frac{2^{2r}}{\sqrt{\pi r}} e^{\frac{1}{9r}} \leqslant \binom{2r}{r} \leqslant \frac{2^{2r}}{\sqrt{\pi r}} e^{\frac{1}{8r}}.$$

PROOF. By using Stirling's approximation [**Rob55**]

$$\sqrt{2\pi r}\left(\frac{r}{e}\right)^r e^{\frac{1}{12r+1}} \leqslant r! \leqslant \sqrt{2\pi r}\left(\frac{r}{e}\right)^r e^{\frac{1}{12r}},$$

we have

$$\begin{aligned}
\binom{2r}{r} = \frac{(2r)!}{(r!)^2} &\geqslant \frac{\sqrt{2\pi 2r}\left(\frac{2r}{e}\right)^{2r} e^{\frac{1}{12r+1}}}{\left(\sqrt{2\pi r}\left(\frac{r}{e}\right)^r e^{\frac{1}{12r}}\right)^2} \\
&= \frac{\sqrt{4\pi r}\left(\frac{2r}{e}\right)^{2r} e^{\frac{2}{12r+1}}}{2\pi r\left(\frac{r}{e}\right)^{2r} e^{\frac{1}{24r}}} \\
&= \frac{2^{2r}}{\sqrt{\pi r}} e^{\frac{2}{12r+1} - \frac{1}{24r}} \geqslant \frac{2^{2r}}{\sqrt{\pi r}} e^{\frac{1}{9r}}.
\end{aligned}$$

The proof of the upper bound is analogous (swap $e^{\frac{1}{12r+1}}$ and $e^{\frac{1}{12r}}$ in the first inequality). ∎

CHAPTER 9

Self-Stabilizing Consensus

In this chapter we prove the results presented in Section 2.6.

As in chapters 5 and 6, we focus on the basic $\mathcal{PULL}$ model of communication, in which in each round, each agent extracts information from few randomly chosen agents. We seek to identify the smallest amount of information revealed in each interaction (message size) that nevertheless allows for efficient and robust computations of fundamental information dissemination tasks. We focus on the *majority bit dissemination* problem that considers a population of n agents, with a designated subset of *source agents*. Each source agent holds an *input bit* and each agent holds an *output bit*. The goal is to let all agents converge their output bits on the most frequent input bit of the sources (the *majority bit*). Note that the particular case of a single source agent corresponds to the classical problem of *broadcast* (also termed *bit dissemination*). We concentrate on the severe fault-tolerant context of *self-stabilization*, in which a correct configuration must be reached eventually, despite all agents starting the execution with arbitrary initial states. In particular, the specification of who is a source and what is its initial input bit may be set by an adversary.

We first design a general compiler which can essentially transform any self-stabilizing algorithm with a certain property (called "the *bitwise-independence property*") that uses ℓ-bits messages to one that uses only $\log \ell$-bits messages, while paying only a small penalty in the running time. By applying this compiler recursively we then obtain a self-stabilizing *clock-synchronization* protocol, in which agents synchronize their clocks modulo some given integer T, within $\tilde{\mathcal{O}}(\log n \log T)$ rounds w.h.p., and using messages that contain 3 bits only.

We then employ the new clock synchronization tool to obtain a self-stabilizing majority bit dissemination protocol which converges in $\tilde{\mathcal{O}}(\log n)$ time, w.h.p., on every initial configuration, provided that the ratio of sources supporting the minority opinion is bounded away from half. Moreover, this protocol also uses only 3 bits per interaction.

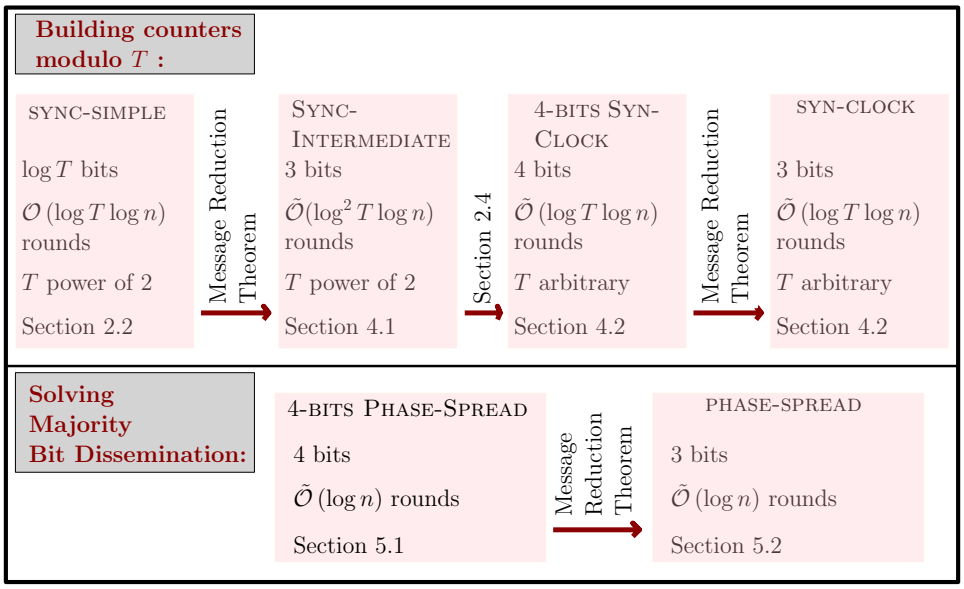

FIGURE 26. The structure of our arguments. Note that the Message Reduction Theorem is used on three occasions.

9.0.1. A majority based, self-stabilizing protocol for consensus on one bit

Let us recall[1] the stabilizing consensus dynamics by Doerr et al. in [**DGM+11**], the 3-Median dynamics. At the outset, each agent holds an opinion. At each round each agent looks at the opinions of two other random agents and updates her opinion taking the majority among the bits of the observed agents and her own. Note that, in the binary-opinion case, this dynamics uses only a single bit per interaction, namely, the node's opinion. The usefulness of 3-Median dynamics comes from its extremely fast and fault-tolerant convergence toward an agreement among agents, as given by the following result.

THEOREM 24 (3-Median dynamics ([**DF11**])). *From any initial configuration, 3-Median dynamics converges to a state in which all agents agree on the same output bit in $\mathcal{O}(\log n)$ rounds, w.h.p. Moreover, if there are at most $\kappa \leqslant n^{1/2-\varepsilon}$ Byzantine agents, for any constant $\varepsilon > 0$, then after $\mathcal{O}(\log n)$ rounds all non-Byzantine agents have converged and consensus is maintained for $n^{\Omega(1)}$ rounds, w.h.p.*

REMARK 7. The original statement of [**DGM+11**] says that if at most $\kappa \leqslant \sqrt{n}$ agents can be corrupted at any round, then convergence happens for all but at most $\mathcal{O}(\kappa)$ agents. Let us explain why we can replace $\mathcal{O}(\kappa)$

[1]The protocols we analyse use this dynamics as a *black box*. However, we note that the constructions we outline are in fact independent of the choice of consensus protocol, and this protocol could be replaced by other protocols that achieve similar guarantees.

by κ, if $\kappa \leqslant n^{\frac{1}{2}-\varepsilon}$. Assume that we are in the regime $\kappa \leqslant n^{\frac{1}{2}-\varepsilon}$. It follows from [**DGM$^+$11**] that all but a set of $\mathcal{O}(\kappa)$ agents reach consensus after $\mathcal{O}(\log n)$ round. This set of size $\mathcal{O}(\kappa)$ contains both Byzantine and non Byzantine agents. However, if the number of agents holding the minority opinion is $\mathcal{O}(\kappa) = \mathcal{O}(n^{1/2-\varepsilon})$, then the expected number of non Byzantine agents that disagree with the majority *at the next round* is in expectation $\mathcal{O}(\kappa^2/n) = \mathcal{O}(n^{-2\varepsilon})$. Thus, by Markov's inequality, this implies, that at the next round consensus is reached among *all non-Byzantine agents*, w.h.p. Note also that, for the same reasons, the Byzantine agents do not affect any other non-Byzantine agent for n^{ε} rounds, w.h.p.

9.0.2. The Setting

9.0.2.1. *The communication model.* As in Chapters 5 and Chapter 6, we adopt the uniform $\mathcal{PULL}$ model [**DGH$^+$88**]. Because of the focus of this chapter in minimizing the message-size, we make explicit two parameters of the model: the number η of agents that each agent observes at each round, and the number ℓ of bits that each agent "displays" and that are visible to other agents. More formally, in the uniform $\mathcal{PULL}(\eta)$ model, communication proceeds in discrete rounds. In each round, each agent u "observes" η arbitrary other agents, chosen u.a.r.among all agents, including herself. To simplify notation, we often omit the parameter η when it is equal to 2.

When an agent u "observes" another agent v, she can peek into a designated *visible part* of v's memory. If several agents observe an agent v at the same round then they all see the same visible part. The *message size* denotes the number of bits stored in the visible part of an agent. Occasionally, we denote with $\mathcal{PULL}(\eta, \ell)$ the $\mathcal{PULL}(\eta)$ model with message size ℓ. We are primarily interested in message size that is independent of n, the number of agents.

9.0.2.2. *Agents.* As in previous chapters, we assume that agents do not have unique identities, that is, the system is *anonymous*. We do not aim to minimize the (non-visible) memory requirement of the agent, yet, we note that our constructions can be implemented with relatively short memory, using $\mathcal{O}(\log \log n)$ bits. We assume that each agent internally stores a clock modulo some integer T, which is incremented at every round. We point out in advance that, in the bit dissemination problem, we set $T = \mathcal{O}(\log n)$.

9.0.2.3. *Majority bit dissemination problem.* We assume a system of n agents each having an internal state that contains an *indicator bit* which indicates whether or not the agent is a *source*. Each source holds a binary *input bit*[2]and each agent (including sources) holds a binary *opinion*. The number of sources (i.e., agents whose indicator bit is 1) is denoted by k. We denote

[2] Note that having the indicator bit equal to 1 is equivalent to possessing an input bit: both are exclusive properties of source nodes. However, we keep them distinct for a clearer presentation.

by k_0 and k_1 the number of sources whose input bit is initially set to 1 and 0, respectively. Assuming $k_1 \neq k_0$, we define the *majority bit*, termed b_{maj}, as 1 if $k_1 > k_0$ and 0 if $k_1 < k_0$. Source agents know that they are sources (using the indicator bit) but they do not know whether they hold the majority bit. The parameters k, k_1 or k_0 are not known to the sources or to any other agent.

It is required that the opinions of all agents eventually converge to the majority bit[3] b_{maj}. We note that agents hold their output and indicator bits privately, and we do not require them to necessarily reveal these bits publicly (in their visible parts) unless they wish to. To avoid dealing with the cases where the number of sources holding the majority bit is arbitrarily close to $\frac{k}{2}$, we shall guarantee correctness (w.h.p.) only if the fraction of sources holding the majority is bounded away from $\frac{1}{2}$, i.e., only if $|\frac{k_1}{k_0} - 1| > \varepsilon$, for some positive constant ε. When $k = 1$, the problem is called *bit dissemination*, for short. Note that in this case, the single source agent holds the bit b_{maj} to be disseminated and there is no other source agent introducing a conflicting opinion.

9.0.2.4. *T-clock synchronization.* Let T be an integer. In the *T-clock synchronization* problem, each agent maintains a *clock* modulo T that is incremented at each round. The goal of agents is to converge on having the same value in their clocks modulo T. (We may omit the parameter T when it is clear from the context.)

9.0.2.5. *Probabilistic self-stabilization and convergence.* Self-stabilizing protocols are meant to guarantee that the system eventually converges to a *legal* configuration regardless of the initial states of the agents [**Dij74**]. Here we adopt the notion of *probabilistic self-stabilization* adopted in Chapter 7 (Definition 9), where stability (closure) is guaranteed only w.h.p. More formally, for the clock synchronization and majority bit dissemination problems, we assume that *all* states are initially set by an adversary except that

- for both problems, it is assumed that the agents know their total number n, and
- for the clock synchronization problems, it is assumed that the agents know the modulo T of the clock that they have to synchronize,

and that these information are not corrupted.

In the context of T-clock synchronization, a legal configuration is reached when all clocks show the same time modulo T, and in the majority bit dissemination problem, a legal configuration is reached when all agents output the majority bit b_{maj}. Note that in the context of the majority bit dissemination problem, the legality criteria depends on the initial configuration (that may be set by an adversary). That is, the agents must converge their

[3]The majority is not defined if $k_1 = k_0$; in this case the only requirement is consensus, i.e. that the outputs of the agents are eventually equal.

opinion on the majority of input bits of sources, as evident in the initial configuration.

Recall that a system is said to *stabilize* in t rounds if, from any initial configuration, within t rounds it reaches a legal configuration and remains legal for at least some polynomial time [**DGM+11**], w.h.p. In fact, for the self-stabilizing bit dissemination problem, if there are no conflicting source agents holding a minority opinion (such as in the case of a single source agent), then our protocols guarantee that once a legal configuration is reached, it remains legal indefinitely. Note that, for any of the problems, we do not require that each agent irrevocably commits to a final opinion but that eventually agents arrive at a legal configuration without necessarily being aware of that.

9.0.3. Protocol Syn-Simple: A simple protocol with many bits per interaction

We now present a simple self-stabilizing T-Clock Synchronization protocol, called Syn-Simple, that uses relatively many bits per message, and relies on the assumption that T is a power of 2. The protocol is based on iteratively applying a self-stabilizing consensus protocol on each bit of the clock separately, and in parallel.

Formally, each agent u maintains a clock $C_u \in [0, T-1]$. At each round, u displays the opinion of her clock C_u, pulls 2 uniform other such clock opinions, and updates her clock as the bitwise majority of the two clocks it pulled, and her own. Subsequently, the clock C_u is incremented. We present the pseudocode of Syn-Simple in Algorithm 3.

Syn-Simple protocol

1: u samples two agents u_1 and u_2.
2: u updates its clock with the bitwise majority of its clock and those of the sample nodes.
3: u increments its clock by one unit.

Algorithm 3. One round of Syn-Simple, executed by each agent u.

We prove the correctness of Syn-Simple in the next proposition.

Proposition 2. *Let T be a power of 2. The protocol* Syn-Simple *is a self-stabilizing protocol that uses $\mathcal{O}(\log T)$ bits per interaction and synchronizes clocks modulo T in $\mathcal{O}(\log T \log n)$ rounds, w.h.p.*

Proof. Let us look at the least significant bit. One round of Syn-Simple is equivalent to one round of 3-Median dynamics with an extra flipping of the opinion due to the increment of the clock. The crucial point is that all agents jointly flip their bit on every round. Because the function agents apply, mode(), is symmetric, it commutes with the flipping operation. More formally, let $\vec{b}_i$ be the vector of the first bits of the clocks of the agents at round i under an execution of Syn-Simple: in other words, $(\vec{b}_i)_u$ is the less

significant bit of node u's clock at time i. We also denote by $\vec{c}_i$ the first bits of the clocks of the agents at round i obtained by running a modified version of SYN-SIMPLE in which *time is not incremented*. (i.e. we skip line 9.0.3 in Algorithm 3). We couple $\vec{b}$ and $\vec{c}$ trivially, by running the two versions on the same interaction pattern (in other words, each agent starts with the same memory and pulls the same agents at each round in both executions). Then, $\vec{b_i}$ is equal to $\vec{c_i}$ on even rounds, while is equal to $\vec{c_i}$ flipped on odd rounds. Moreover, we know from Theorem 24 that $\vec{c_i}$ converge to a stable opinion in a self-stabilizing manner. It follows that, from any initial configuration of states (i.e. clocks), after $\mathcal{O}(\log n)$ rounds of executing SYN-SIMPLE, all agents share the same opinion for their first bit, w.h.p, and jointly flip it in each round. Once agents agree on the first bit, since T is a power of 2, the increment of time makes them flip the second bit *jointly* once every 2 rounds[4]. More generally, assuming agents agree on the first ℓ bits of their clocks, they *jointly* flip the $\ell+1$'st bit once every 2^ℓ rounds, on top of doing the 3-Median dynamics protocol on that bit. Therefore, the same coupling argument shows that the flipping doesn't affect the convergence on bit $\ell+1$. Thus, $\mathcal{O}(\log n)$ rounds after the first ℓ bits are synchronized, the $\ell+1$'st bit is synchronized as well w.h.p. The result thus follows by induction. ∎

9.0.4. The bitwise-independence property

In Section 9.1, we describe a general transformer which is useful for reducing the message size of protocols with a certain property called *bitwise-independence*. Before defining the property we need to define a variant of the $\mathcal{PULL}$ model, which we refer to as the $\mathcal{BIT}$ model. The reason we introduce such a variant is mainly technical, as it appears naturally in the proofs.

Recall that in the $\mathcal{PULL}(\eta, \ell)$ model, at any given round, each agent u is reading an ℓ-bit message m_{v_j} for each of the η observed agents v_j chosen u.a.r. (in our case $\eta = 2$), and then, in turn, u updates her state according to the instructions of a protocol Ψ. Informally, in the $\mathcal{BIT}$ model, each agent u also receives η messages, however, in contrast to the $\mathcal{PULL}$ model where each such message corresponds to one observed agent, in the $\mathcal{BIT}$ model, the i'th bit of each such message is received independently from an agent, chosen u.a.r. from all agents.

DEFINITION 23 (The $\mathcal{BIT}$ model). In the $\mathcal{BIT}$ model, at each round, each agent u picks $\eta\ell$ agents u.a.r., namely,

$$v_1^{(1)}, v_2^{(1)}, \dots v_\ell^{(1)}, \dots, v_1^{(\eta)}, v_2^{(\eta)}, \dots v_\ell^{(\eta)},$$

and reads $\hat{s_i}^{(j)} = s_i(v_i^{(j)})$, the i-th bit of the visible part of agent $v_i^{(j)}$, for every $i \leqslant \ell$ and $j \leqslant \eta$. For each $j \leqslant \eta$, let $\hat{m_j}(u)$ be the ℓ-bit string

$$\hat{m_j}(u) := (\hat{s_1}^{(j)}, \hat{s_2}^{(j)}, \dots, \hat{s_\ell}^{(j)}).$$

[4]To get the feeling of the kind of dependence more significant bits have on the less significant ones when T is not a power of 2 observe that, for example, if $T = 3$ then the first bit takes cyclically the values 1, 0 and again 0.

By a slight abuse of language we call the strings $\{\hat{m_j}(u)\}_{j\leqslant\eta}$ the *messages* received by u in the $\mathcal{BIT}$ model.

We are now ready to define the special property that we have mentioned above.

DEFINITION 24 (The *bitwise−independence* property). Consider a protocol Ψ designed to work in the $\mathcal{PULL}$ model. We say that Ψ has the bitwise-independence property if its correctness and running time guarantees remain the same, under the $\mathcal{BIT}$ model, assuming that given the messages $\{\hat{m_j}(u)\}_{j\leqslant\eta}$ it receives at any round, each agent u performs the same actions that it would have, had it received these messages in the $\mathcal{PULL}$ model.

Let us first state a fact about protocols having the bitwise-independence property.

LEMMA 70. *Assume protocol* SYN-GENERIC *is a protocol synchronizing clocks modulo T for some T and protocol* P *is a protocol which works assuming agents share a clock modulo T. Denote by* SYN-P *the parallel execution of* SYN-GENERIC *and* P*, with* P *using the clock synchronized by* SYN-GENERIC*. If* SYN-GENERIC *and* P *are self-stabilizing then so is* SYN-P*, and the convergence time of* SYN-P *is at most the sum of convergence times of* SYN-GENERIC *and* P*. Finally, if* SYN-GENERIC *and* P *have the bitwise-independence property, and* P *is also self-stabilizing,* SYN-P *has the bitwise-independence property too.*

PROOF. Since SYN-P consists in the parallel execution of SYN-GENERIC and P, the configuration of the system $\mathbf{C}^{(t)}_{\text{SYN-P}} = (\mathbf{C}^{(t)}_{\text{SYN-GENERIC}}, \mathbf{C}^{(t)}_{\text{P}})$ at time t is composed by a a first part $\mathbf{C}^{(t)}_{\text{SYN-GENERIC}}$ which describes the nodes' state as for the execution of SYN-GENERIC, and a second part $\mathbf{C}^{(t)}_{\text{P}}$ which describes the nodes' state as for the execution of P.

Let $T_{\text{SYN-GENERIC}}$ and T_{P} be upper bounds on the convergence time of SYN-GENERIC and P, respectively.

As for the self-stabilizing property of SYN-P and its convergence time, observe that since SYN-GENERIC is self-stabilizing, there exist a time $t_1 \leqslant T_{\text{SYN-GENERIC}}$ such that $\mathbf{C}^{(t_1)}_{\text{SYN-GENERIC}}$ is legitimate, i.e. such that the nodes' clocks modulo T are synchronized, and by definition of self-stabilization (closure property), they remain synchronized from that moment on. Then, since P is self-stabilizing as well, no matter what $\mathbf{C}^{(t_1)}_{\text{P}}$ is: there exist a time $t_2 \leqslant t_1 + T_{\text{P}} \leqslant T_{\text{SYN-GENERIC}} + T_{\text{P}}$ such that $\mathbf{C}^{(t_1)}_{\text{P}}$ is legitimate, i.e. such that P has converged, which also means that SYN-P correctly converges in at most $T_{\text{SYN-GENERIC}} + T_{\text{P}}$ rounds.

As for the bitwise-independence property, assume we run SYN-P in the $\mathcal{BIT}$ model. The execution of SYN-GENERIC is carried independently of the execution of P. Since, by hypothesis, SYN-GENERIC has the independence property, eventually all agents have a synchronized clock modulo T. Thus,

once clocks are synchronized, we can disregard the part of the message corresponding to SYN-GENERIC, and view the execution of SYN-P as simply P. Therefore, since P is self-stabilizing and has the independence property, SYN-P still works in the $\mathcal{BIT}$ model as in the original $\mathcal{PULL}$ model. ⊠

We next show that the protocol SYN-SIMPLE has the aforementioned bitwise-independence property.

LEMMA 71. SYN-SIMPLE *has the bitwise-independence property.*

PROOF. Let ℓ' be the size of the clocks. Assume the first $i < \ell'$ bits of the clocks have been synchronized. At this stage, the $(i+1)$-st bit of each agent u is flipped every 2^i rounds and updated as the majority of the $(i+1)$-st bit of $C(u)$ and the 2 pulled messages on each round. Since the first ℓ' bits are synchronized, the previous flipping is performed by all agents at the same round. The thesis follows from the observation that, in order for SYN-SIMPLE to work, we do not need the bit at index $(i+1)$ to come from the same agent as those bits used to synchronize the other indices, as long as convergence on the first i bits has been achieved. ⊠

9.1. A General Compiler that Reduces Message Size

In this section we present a general compiler that allows to implement a protocol Ψ using ℓ-bit messages while using messages of order $\log \ell$ instead, as long as Ψ enjoys the bitwise-independence property. The compiler is based on replacing a message by an index to a (dynamic) bit of the message. This tool is repeatedly used in the following sections to obtain our clock synchronization and majority bit dissemination algorithms that use 3-bit messages.

THEOREM 17 (Message Reduction Theorem). *Any self-stabilizing protocol Ψ in the $\mathcal{PULL}(\eta, \ell)$ model having the bitwise-independence property, and whose running time is L_Ψ, can be emulated by a protocol* EMUL(Ψ) *which runs in*[5] *the $\mathcal{PULL}(2, \lceil \log(\frac{\eta}{2}\ell) \rceil + 1)$ model, has running time $\mathcal{O}(\log(\eta\ell)\log n + \frac{\eta}{2}\ell L_\Psi)$ and has itself the bitwise-independence property.*

PROOF OF THEOREM 17. Let $s(u) \in \{0,1\}^\ell$ be the message displayed by an agent u under Ψ at a given round. For simplicity's sake, in the following we assume that η is even, the other case is handled similarly. In EMUL(Ψ),

[5] The only reason for designing EMUL(Ψ) to run in the

$$\mathcal{PULL}\left(2, \lceil \log\left(\frac{\eta}{2}\ell\right) \rceil + 1\right)$$

model in the Message Reduction Theorem is the consensus protocol we adopt, 3-Median dynamics, which works in the $\mathcal{PULL}(2)$ model. In fact, EMUL(Ψ) can be adapted to run in the

$$\mathcal{PULL}\left(1, \lceil \log\left(\eta\ell\right) \rceil + 1\right)$$

model by using a consensus protocol which works in the $\mathcal{PULL}(1)$ model. However, no self-stabilizing binary consensus protocol in the $\mathcal{PULL}(1)$ model with the same performances as 3-Median dynamics is currently known.

agent u keeps the message $s(u)$ privately, and instead displays a clock $C(u)$ written on $\lceil \log(\frac{\eta}{2}\ell) \rceil$ bits, and one bit of the message $s(u)$, which we refer to as the Ψ-bit. Thus, the total number of bits displayed by the agent operating in $\textsc{Emul}(\Psi)$ is $\lceil \log(\frac{\eta}{2}\ell) \rceil + 1$. The purpose of the clock $C(u)$ is to indicate to agent u which bit of $s(u)$ to display. In particular, if the counter has value 0, then the 0-th bit (i.e the least significant bit) of $s(u)$ is shown as the Ψ-bit, and so on. In what follows, we refer to $s(u)$ as the *private message* of u, to emphasize the fact that this message is not visible in $\textsc{Emul}(\Psi)$. See Figure 27 for an illustration.

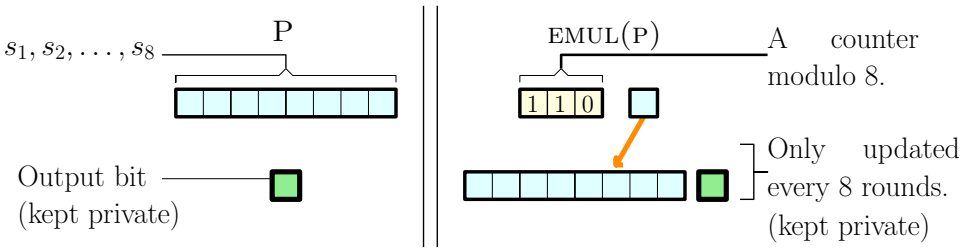

FIGURE 27. On the left is a protocol Ψ using $\ell = 8$ bits in total and pulling only one node per round ($\eta = 1$). On the right is the emulated version $\textsc{Emul}(\Psi)$ which uses 4 bits only. The bits depicted on the bottom of each panel are kept privately, while the bits on the top are public, that is, appear in the visible part.

Each round of Ψ executed in the $\mathcal{PULL}(\eta, \ell)$ model by an agent u is emulated by $\frac{\eta}{2}\ell$ rounds of $\textsc{Emul}(\Psi)$ in the $\mathcal{PULL}(2, \lceil \log(\frac{\eta}{2}\ell) \rceil + 1)$ model. We refer to such $\frac{\eta}{2}\ell$ rounds as a *phase*, which is further divided to $\frac{\eta}{2}$ subphases of length ℓ. Note that since each agent samples 2 agents in a round, the total number of agents sampled by an agent during a phases is $\eta\ell$.

For a generic agent u, a phase starts when its clock $C(u)$ is zero, and ends after a full loop of its clock (i.e. when $C(u)$ returns to zero). Each agent u is running protocol SYN-SIMPLE on the $\lceil \log(\frac{\eta}{2}\ell) \rceil$ bits which correspond to her clock $C(u)$. Note that the phases executed by different agents may initially be unsynchronized, but, thanks to Proposition 2, the clocks $C(u)$ eventually converge to the same value, for each agent u, and hence all agents eventually agree on when each phase (and subphase) starts.

Let u be an arbitrary agent. Denote by

$$\hat{s}_1^{(1)}, \hat{s}_2^{(1)}, \ldots \hat{s}_\ell^{(1)}, ..., \hat{s}_1^{(\eta)}, \hat{s}_2^{(\eta)}, \ldots \hat{s}_\ell^{(\eta)}$$

the Ψ-bits collected by u from agents chosen u.a.r during a phase. Consider a phase and a round $z \in \{1, \cdots, \frac{\eta}{2}\ell\}$ in that phase. Let i and j be such that $z = j \cdot \ell + i$. We view z as round i of subphase $j + 1$ of the phase. On this round, agent u pulls two messages from agents v and w, chosen u.a.r. Once the clocks (and thus phases and subphases) have synchronized, agents v and

w are guaranteed to be displaying the ith index of their private messages, namely, the values $s_i(v)$ and $s_i(w)$, respectively. Agent u then sets $\hat{s}_i^{(2j-1)}$ equal to $s_i(v)$ and $\hat{s}_i^{(2j)}$ equal to $s_i(w)$.

In EMUL(Ψ), the messages displayed by agents are only updated after a full loop of C. It therefore follows from the previous paragraph that the Ψ-bits collected by agent u after a full-phase are distributed like the bits collected during one round of Ψ in the $\mathcal{BIT}$ model (see Definition 23), assuming the clocks are already synchronized.

Correctness. The bitwise-independence property of SYN-SIMPLE (Lemma 71), implies that SYN-SIMPLE still works when messages are constructed from the Ψ-bits collected by EMUL(Ψ). Therefore, from Proposition 2, eventually all the clocks C are synchronized. Since private messages s are only updated after a full loop of C, once the clocks C are synchronized a phase of EMUL(Ψ) corresponds to *one* round of Ψ, executed in the $\mathcal{BIT}$ model. Hence, the hypothesis that Ψ operates correctly in a self-stabilizing way in the $\mathcal{BIT}$ model implies the correctness of EMUL(Ψ).

Running time. Once the clocks $C(u)$ are synchronized, for all agents u, using the first $\lceil \log(\frac{\eta}{2}\ell) \rceil$ bits of the messages, the agents reproduce an execution of Ψ with a multiplicative time-overhead of $\frac{\eta}{2}\ell$. Moreover, from Proposition 2, synchronizing the clocks $C(u)$ takes $\mathcal{O}(\log(\eta m)\log n)$ rounds. Thus, the time to synchronize the clocks costs only an additive factor of $\mathcal{O}(\log(\eta m)\log n)$ rounds, and the total running time is $\mathcal{O}(\log(\eta m)\log n) + \frac{\eta}{2}\ell \cdot L_\Psi$.

Bitwise-independence property. Protocol EMUL(Ψ) inherits the bitwise-independence property from that of SYN-SIMPLE (Lemma 71) and Ψ (which has the property by hypothesis): We can apply Lemma 70 where SYN-GENERIC is SYN-SIMPLE and P is the subroutine described above, which displays at each round the bit of Ψ whose index is given by a synchronized clock C modulo ℓ (i.e. the one produced by SYN-SIMPLE). Observe that the aforementioned subroutine is self-stabilizing, since it emulates Ψ once clocks are synchronized. Then, in the notation of Lemma 70, EMUL(Ψ) is SYN-P. ∎

9.2. Self-Stabilizing Clock Synchronization

SYN-INTERMEDIATE protocol

MEMORY: Each agent u keeps a sequence of clocks $C_1, \ldots, C_\tau$ and a sequence of bits $b_1, \ldots, b_\tau$. The clock C_1 runs modulo T, the clock C_τ runs modulo 4, and the i-th clock C_i runs modulo $2^{\ell_i - 1}$ (see proof of Lemma 72). Each agent u also maintains a sequence of heaps (or some ordered structure) S_i^δ, for each $\delta \in \{1, 2\}$ and $i = 1, \ldots, \tau$.

MESSAGE: u displays C_τ (2 bits) and b_τ (1 bit). For all $i \in [\tau],, b_i(u)$ is the $C_i(u)$-th bit of the string obtained concatenating the binary representation of $C_{i-1}(u)$ and $b_{i-1}(u)$.

1: u samples two agents u_1 and u_2.
2: u updates its clock with the bitwise majority of its clock and those of the sample nodes.
3: u increments its clock by one unit.
4: u sets i^* equal to the maximal $i < \tau$ such that $C_{i+1} \neq 0$.
5: For $\delta = 1, 2$, u pushes $b_\tau(u_\delta)$ in $S_{i^*}^\delta$.
(Note that, if $C_{i^*+1}, \ldots, C_\tau$ are synchronized, then all agents are displaying the bit with index C_{i^*+1} of (C_{i^*}, b_{i^*}) as b_τ.)
6: While $i > 1$ and $C_i = 0$, u does the following:
7: | Pops the last $m_{i-1} - 1$ bits from S_{i-1}^δ and set s^δ equal to it.
8: | Sets C_{i-1} equal to the bitwise majority of $C_{i-1}(u)$, s^1 and s^2.
9: | Increments C_{i-1} and decrement i by one unit.

Algorithm 4. Iterative version of the protocol SYN-INTERMEDIATE, executed by each agent u, unfolding the recursion in proof of Lemma 72.

In Section 9.0.3 we described SYN-SIMPLE - a simple self-stabilizing clock synchronization protocol that uses $\log T$ bits per interaction. In this section we describe our main self-stabilizing clock synchronization protocol, SYN-3BITS, that uses only 3 bits per interaction. We first assume T is a power of 2. We show how to get rid of this assumption in Section 9.2.2.

9.2.1. Clock Synchronization with 3-bit messages, assuming T is a power of two

In this section, we show the following result.

LEMMA 72. *Let T be a power of 2. There exists a synchronization protocol* SYN-INTERMEDIATE *which synchronizes clocks modulo T in time $\tilde{\mathcal{O}}\left(\log^2 T \log n\right)$ using only 3-bit messages. Moreover,* SYN-INTERMEDIATE *has the bitwise-independence property.*

Before presenting the proof of Lemma 72, we need a remark about clocks.

REMARK 8. In order to synchronize a clock C modulo T, throughout the analysis we often obtain a clock C' modulo T *which is incremented every* ℓ *rounds*. However, C' can still be translated back to a clock modulo T which is incremented every round, by keeping a third clock C'' modulo ℓ and setting

$$C = C' + C'' \mod T.$$

PROOF OF LEMMA 72. At a high level, we simply apply iteratively the Message Reduction Theorem in order to reduce the message to 3 bits, starting with $\Psi =$ SYN-SIMPLE. A pictorial representation of our recursive protocol is given in Figure 28, and a pseudocode is given in Algorithm 4[6]

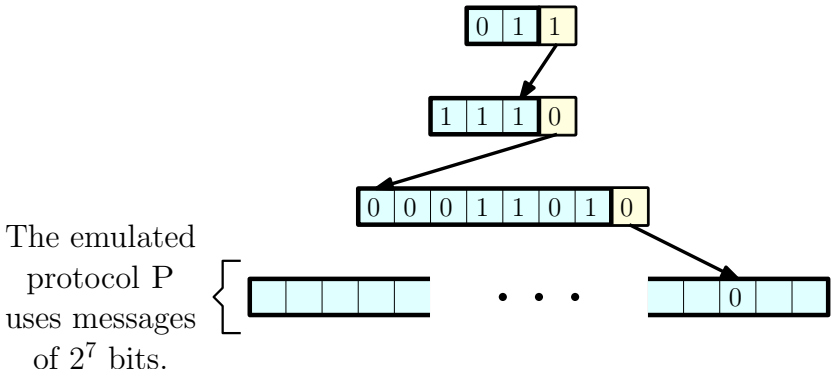

FIGURE 28. A more explicit view of our 3-bit emulation of protocol Ψ, obtained by iterating Lemma 17. The down-most layer represents the 2^7-bits message displayed by protocol Ψ. Each layer on the picture may be seen as the message of a protocol emulating Ψ with fewer bits, that is, as we go up on the figure we obtain more and more economical protocols in terms of message length. In particular, the top layer represents the 3-bit message in the final emulation. The left-most part of each message (colored in light blue) encodes a clock. The right-most bit (colored in light yellow) of each message (except the bottom-most one) corresponds to a particular bit of the layer *below* it. The index of this particular displayed bit is given by the value of the clock. Each clock on an intermediate layer is updated only when the clock on the layer *above* completes a loop (i.e., has value 0). The clock on the top-most layer is updated on every round.

Let us consider what we obtain after applying the Message Reduction Theorem the first time to $\Psi =$SYN-SIMPLE for clocks modulo T. Recall that we assume that T is a power of 2. From Proposition 2 we know that

[6]The pseudocode deviates from the presentation done in the proof, as it makes no use of recursion.

in this case, the convergence time of Syn-Simple is $L_\Psi = \mathcal{O}(\log T \log n)$, the number of pulled agents at each round is $m = 2$ and the number of bits of each message is $\ell = \log T$. Note that, with the emulation produced by the Message Reduction Theorem, the clock used in $\Psi =$Syn-Simple is incremented only every $\ell = \log T$ rounds. By Remark 8, we can translate the latter clock to the desired clock modulo T (which is incremented at each round).

Hence, by the running time analysis of the Message Reduction Theorem, we obtain a protocol Emul(Ψ) which synchronizes a clock modulo T in

$$\mathcal{O}(\log n \log\log T) + \mathcal{O}\left(\log^2 T \log n\right) = \mathcal{O}\left(\log^2 T \log n\right)$$

rounds. The message size is reduced from $\log T$ to

$$\lceil \log\log T \rceil + 1 = \mathcal{O}(\log\log T)\,.$$

By repeatedly applying the Message Reduction Theorem, we reduce the size of the message ℓ as long as $\ell > \lceil \log \ell \rceil + 1$, i.e. as long as $\ell > 3$. The number of repeated application of the Message Reduction Theorem until the message size is 3 is thus of order $\log^* T$.

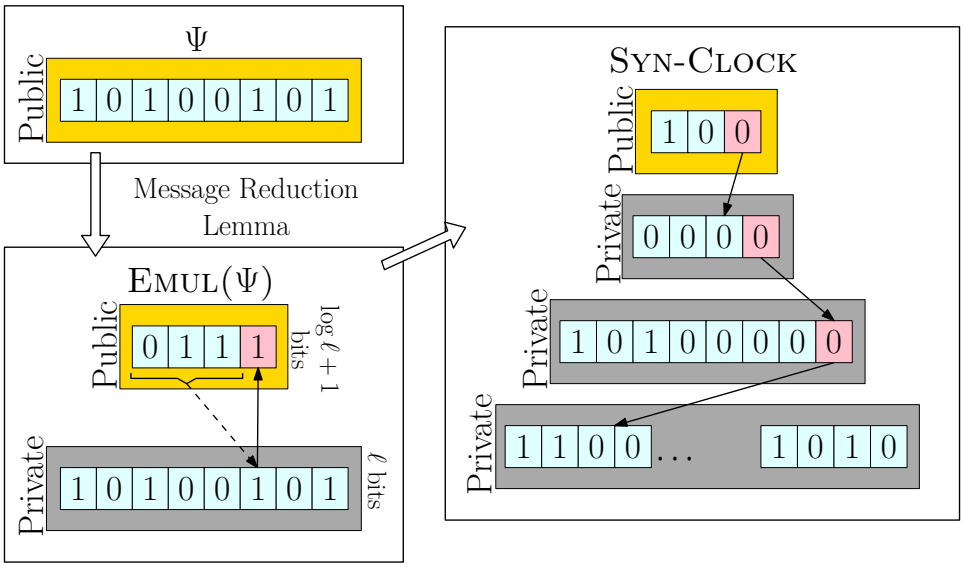

Figure 29. Illustration which combines figures 27 and 28. The yellow boxes represent the visible part of agents' memory.

Let us analyze the running time. Let $\ell_1 = \log T$, $\ell_{i+1} = \lceil \log \ell_i \rceil + 1$ and let $\tau(T) = \tau$ be the smallest integer such that $\ell_\tau = 3$. We apply the Message Reduction Theorem $i \leqslant \tau$ times, and we obtain a message size ℓ_i and a running time L_i, such that

$$L_{i+1} \leqslant \gamma_1(\log \ell_i \log n + \ell_i L_i), \tag{167}$$

for some constant γ_1 independent of i. We set L_1 to be

$$L_1 := L_{\text{Syn-Simple}} \vee \log n = \mathcal{O}\left(\log T \log n\right) \vee \log n,$$

taking the maximum with $\log n$ for technical convenience. The second term dominates in (167) because $\ell_i >> \log \ell_i$ and $L_i > \log n$. Hence L_i is at most of order $\prod_{j<i} \ell_j \cdot L_1$. More precisely, by induction we can bound $L_i \leqslant \gamma_1^i \prod_{j=1}^{i-1} \ell_j L_1$, since

$$\begin{aligned} L_{i+1} &\leqslant \gamma_1 \log \ell_i \log n + \gamma_1^i \prod_{j=1}^{i} \ell_j \cdot L_1 \\ &\leqslant \gamma_1 \ell_i \log n + \gamma_1^i \prod_{j=1}^{i} \ell_j \cdot L_1 \\ &\leqslant 2\gamma_1^i \prod_{j=1}^{i} \ell_j \cdot L_1 \leqslant \gamma_1^{i+1} \prod_{j=1}^{i} \ell_j \cdot L_1, \end{aligned}$$

where we use the fact that $\gamma_1 > 2$, and the definition of L_1.

The running time of $\textsc{Emul}(\Psi)$ =Syn-Clock after the last application of the Message Reduction Theorem, i.e. τ, is thus

$$L_{\text{Syn-Clock}} := L_\tau \leqslant \gamma_1^\tau \prod_{i=1}^{\tau} \ell_i L_1.$$

We now use the bounds

$$\begin{aligned} L_1 &= \mathcal{O}(\log T \log n), \\ \prod_{i=1}^{\tau} \ell_i &\leqslant \ell_1 \ell_2 \ell_3^\tau, \\ \ell_1 &= \mathcal{O}\left(\log T\right), \\ \ell_2 &= \mathcal{O}\left(\log \log T\right), \end{aligned}$$

and finally, by Lemma 75 (see in Section 9.4 at the end of the chapter), $\gamma_1^\tau = \mathcal{O}\left(\log \log \log T\right)$ and

$$\ell_3^\tau \leqslant 2^{\mathcal{O}\left((\log^{\circledast 4} T)^2\right)} \leqslant 2^{\mathcal{O}(\log \log \log T)} \leqslant \left(\log \log T\right)^{\mathcal{O}(1)}.$$

We thus conclude that

$$\begin{aligned} L_{\text{Syn-Clock}} &\leqslant \gamma_1^\tau \prod_{i=1}^{\tau} \ell_i L_1 \leqslant \mathcal{O}\left(\log \log \log T\right) \cdot \ell_1 \ell_2 \ell_3^\tau \cdot \mathcal{O}(\log T \log n) \\ &\leqslant \mathcal{O}\left(\log \log \log T\right) \cdot \mathcal{O}\left(\log T\right) \cdot \mathcal{O}\left(\log \log T\right) \\ &\quad \cdot \mathcal{O}\left(\log \log T\right)^{\mathcal{O}(1)} \cdot \mathcal{O}(\log T \log n) \\ &\leqslant \log^2 T \log n \cdot \left(\log \log T\right)^{\mathcal{O}(1)}. \end{aligned}$$

The total slowdown with respect to Syn-Simple corresponds to $\prod_{i=1}^{\tau} \ell_i = \tilde{\mathcal{O}}(\log T)$. Hence the clock produced by the emulation is incremented every $\tilde{\mathcal{O}}(\log T)$ rounds. In other words we obtain a clock modulo $T \cdot f(T)$ for some function f. By Remark 8 we can still view this as a clock modulo T. ▯

9.2.2. Extension to general T and running time improvement

In this subsection we aim to get rid of the assumption that T is a power of 2 in Lemma 72, and also reduce the running time of our protocol to $\tilde{\mathcal{O}}(\log n \log T)$, proving Theorem 16.

Syn-Clock protocol

Memory: Each agent u stores a clock $C'(u)$ which runs modulo $T' \gg \gamma \log n \log T$. Each agent u also stores a variable Q which is incremented only once every T' rounds and runs modulo T.

Message: Each agent u displays 4 bits. On the first 3 bits, protocol Syn-Intermediate is applied to synchronize C'. The 4-th bit $b(u)$ is the bit with index $(\lfloor \frac{C'(u)}{\gamma \log n} \rfloor \mod \lceil \log T \rceil)$ of $Q(u)$.

1: u samples two agents u_1 and u_2.
2: u updates $b(u)$ with the majority of $b(u)$, $b(u_1)$ and $b(u_2)$.
3: If $C' = 0$, increment Q by one unit modulo T.

Output: The clock modulo T is obtained as $C := (C' + Q \cdot T') \mod T$

Algorithm 5. The protocol 4-bit Syn-Clock, executed by each agent u.

Proof of Theorem 16. From Lemma 72, we know that protocol Syn-Intermediate synchronizes clocks modulo T in time $\tilde{\mathcal{O}}\left(\log^2 T \log n\right)$ using only 3-bit messages, provided that T is a power of 2. While protocol Syn-Intermediate emulates protocol Syn-Simple, it displays the first bit of the message of Syn-Simple only once every $\tilde{\mathcal{O}}(\log T)$ rounds. Of course, it would be more efficient to display it $\mathcal{O}(\log n)$ times in a row, so that the 3-Median dynamics would make every agent agree on this bit, and then move to agreeing on the second bit, and so on. To achieve this, as in the proof of Syn-Simple, we can view a clock modulo T, say Q, as written on $\log T$ bits. If agents already possess a "small" counter modulo $T' := \mathcal{O}(\log T \log n)$ they can use it to display the first bit for $\mathcal{O}(\log n)$ rounds, then the second one for $\mathcal{O}(\log n)$ rounds, and so on until each one of the $\lceil \log T \rceil$ bits of T has been synchronized. This would synchronize all bits of the desired clock within $\mathcal{O}(\log T \log n)$ rounds, w.h.p., while being very economical in terms of message length, since only 1 bit is displayed at any time.

Therefore, we can use Lemma 72 to synchronize a counter modulo $\mathcal{O}(\log T \log n)$ in $\tilde{\mathcal{O}}((\log \log T)^2 \log n)$ rounds, using 3 bits per message. Then, we can use a fourth bit to run 3-Median dynamics on each of the $\log T$ bits of Q for $\mathcal{O}(\log n)$ consecutive rounds, for a total running time of $\mathcal{O}(\log T \log n)$

rounds. At this point, an application of the Message Reduction Theorem would give us a protocol with running time $\mathcal{O}(\log T \log n)$ using 3-bit messages. However, perhaps surprisingly, a similar strategy enables us to synchronize a clock modulo any integer (not necessarily a power of 2).

Let us assume that $T \in \mathbb{N}$ is an arbitrary integer. Let $\gamma \log n$ be an upper bound on the convergence time of 3-Median dynamics which guarantees a correct consensus with probability at least $1-n^{-2}$, for some constant γ large enough [**DGM+11**]. Let T' be the smallest power of 2 bigger than

$$\log T \cdot (\gamma \log n + \gamma \log \log T) .$$

By Lemma 72, using 3 bits, the agents can build a synchronized clock C' running modulo T' in time $\tilde{\mathcal{O}}((\log \log T)^2 \log n)$. The other main ingredient in this construction is another clock $Q_{T'}$ which is incremented once every T' rounds and runs modulo T. The desired clock modulo T, which we denote C, is obtained by

$$C := \left(C' + Q_{T'} \cdot T'\right) \mod T.$$

It is easy to check, given the definitions of C' and $Q_{T'}$ that this choice indeed produces a clock modulo T.

It remains to show how the clock $Q_{T'}$ modulo T is synchronized. On a first glance, it may seem as if we did not simplify the problem since Q is a clock modulo T itself. However, the difference between $Q_{T'}$ and a regular clock modulo T is that $Q_{T'}$ is incremented only once every T' rounds. This is exploited as follows.

The counter $Q_{T'}$ is written on $\lceil \log T \rceil$ internal bits. We show how to synchronize $Q_{T'}$ using a 4-th bit in the messages, similarly to the aforementioned strategy to synchronize Q; we later show how to remove this assumption using the Message Reduction Theorem. Let us call a loop of C' modulo T' an *epoch*. The rounds of an epoch are divided in phases of equal length $\gamma \log n + \gamma \log \log T$ (the remaining $T' \mod (\gamma \log n + \gamma \log \log T)$ rounds are just ignored). The clock C' determines which bit from $Q_{T'}$ to display. The first bit of $Q_{T'}$ is displayed during the first phase, then the second one is displayed during the second phase, and so on. By Theorem 24, the length of each phase guarantees that consensus is achieved on each bit of $Q_{T'}$ via[7] 3-Median dynamics, w.h.p. More precisely, after the first bit has been displayed for $\gamma \log n + \gamma \log \log T$ rounds, all agents agree on it with probability[8]

[7]Observe that, once clock C' is synchronized, the bits of $Q_{T'}$ do not change for each agent during each subphase. Thus, we may replace 3-Median dynamics by the Min protocol where on each round of subphase i each agent u pulls another agent v u.a.r. and updates her i-th bit of Q to the minimum between her current i-th bit of Q and the one of v. However, for simplicity's sake, we reuse the already introduced 3-Median dynamics protocol.

[8]From Theorem 24, we have that after $\gamma \log n$ rounds, with γ large enough, the probability that consensus has not been reached is smaller than $\frac{1}{n^2}$. Thus, after $N \cdot \gamma \log n$ rounds, the probability that consensus has not been reached is smaller than $\frac{1}{n^{2N}}$. If we choose $N \log n = \log n + \log \log T$, we thus get the claimed upper bound $\frac{1}{n^2 \log T}$.

$1 - \frac{1}{n^2 \log T}$, provided γ is large enough. Thus, at the end of an epoch, agents agree on all $\lceil \log T \rceil$ bits of $Q_{T'}$ with probability greater than

$$\left(1 - \frac{1}{n^2 \log T}\right)^{\log T} \gg 1 - \mathcal{O}(n^{-2}).$$

We have thus shown that, by the time C' reaches its maximum value of T', i.e. after one epoch, all agents agree on $Q_{T'}$, w.h.p., and then increment it jointly. From Lemma 72, Syn-Intermediate takes

$$\begin{aligned}\tilde{\mathcal{O}}\left(\log^2 T' \log n\right) &= \mathcal{O}\left((\log\log n + \log\log T)^2 \log n\right) \\ &= \mathcal{O}\left(\left((\log\log n)^2 \log n + (\log\log T)^2 \log n\right)\right)\end{aligned}$$

rounds to synchronize a clock C' modulo T', w.h.p. Together with the $\log T\,(\gamma \log n + \gamma \log\log T)$ rounds to agree on $Q_{T'}$, w.h.p., this implies that after

$$\log T \log n \cdot (\log\log T)^{\mathcal{O}(1)} \cdot (\log\log n)^{\mathcal{O}(1)} = \tilde{\mathcal{O}}\left(\log T \log n\right)$$

rounds the clocks C are all synchronized, w.h.p.

Finally, we show how to get rid of the extra 4-th bit to achieve agreement on $Q_{T'}$. Observe that, once C' is synchronized, this bit is used in a self-stabilizing way. Thus, since Syn-Intermediate has the bitwise-independence property, using Lemma 70, the protocol we described above possesses the bitwise-independence property too. By using the Message Reduction Theorem we can thus reduce the message size from 4 bits to $\lceil \log 4 \rceil + 1 = 3$ bits, while only incurring a constant multiplicative loss in the running time. The clock we obtain, counts modulo T but is incremented every 4 rounds only. However, from Remark 8, we can still translate this into a clock modulo T. gg

Remark 9 (Internal memory space). The internal memory space needed to implement our protocols Syn-Simple, Syn-Intermediate, and Syn-Clock is close to $\log T$ in all cases: protocol Syn-Simple uses one counter written on $\log T$ bits, Syn-Intermediate needs internal memory of size

$$\log T + \mathcal{O}\left(\log\log T + \log\log\log T + \ldots\right) \leqslant \log T(1 + o(1)),$$

and the internal memory requirement of Syn-Clock is of order $\log T + \log\log n$.

9.3. Majority Bit Dissemination with a Clock

In this section we assume that agents are equipped with a synchronized clock C modulo $\gamma \log n$ for some big enough constant $\gamma > 0$. In the previous section we showed how to establish such a synchronized clock in $\tilde{\mathcal{O}}(\log n)$ time and using 3-bit messages. We have already seen in Section 2.6.1 how to solve the bit dissemination problem (when we are promised to have a single source agent) assuming such synchronized clocks, by paying an extra bit in the message size and an $\mathcal{O}(\log n)$ additive factor in the running time. This

section is dedicated to showing that, in fact, the more general majority bit dissemination problem can be solved with the same time complexity and using 3-bit messages, proving Theorem 15.

In Section 9.3.1, we describe and analyze protocol Syn-Phase-Spread, which solves majority bit dissemination by paying only a $\mathcal{O}(\log n)$ additive overhead in the running time w.r.t. clock synchronization. For clarity's sake, we first assume that the protocol is using 4 bits (i.e. 1 additional bit over the 3 bits used for clock synchronization), and we later show how to decrease the number of bits back to 3 in Section 9.3.2, by applying the Message Reduction Theorem.

The main idea behind the 3(+1)-bit protocol, called Syn-Phase-Spread, is to make the sources' input bits disseminate on the system in a way that preserves the initial ratio $\frac{k_1}{k_0}$ between the number of sources supporting the majority and minority input bit. This is achieved by dividing the dissemination process in phases, similarly to the main protocol in [**FHK14**] which was designed to solve the bit dissemination problem in a variant of the $\mathcal{PUSH}$ model in which messages are affected by noise. The phases induces a spreading process which allows to leverage on the concentration property of the Chernoff bounds, preserving the aforementioned ratio. While, on an intuitive level, the role of noisy messages in the model considered in [**FHK14**] may be related to the presence of sources having conflicting opinion in our setting, we remark that the protocol presented here and its analysis depart from those of [**FHK14**] on several key points: while the protocol in [**FHK14**] needs to know the the noise parameter, Syn-Phase-Spread does not assume any knowledge about the number of different sources, and the analysis we present does not require to control the growth of the number of speaking agents from above[9].

In order to perform such spreading process with 1 bit only, the protocol in [**FHK14**] leverages on the fact that in the $\mathcal{PUSH}$ model agents can choose *when to speak*, i.e. whether to send a message or not. To emulate this property in the $\mathcal{PULL}$ model, we use the parity of the clock C: on odd rounds agents willing to "send" a 0 display 0, while others display 1 and conversely on even rounds. Rounds are then grouped by two, so 2 rounds in the $\mathcal{PULL}$ model correspond to 1 round in the $\mathcal{PUSH}$ version.

9.3.1. Protocol Syn-Phase-Spread

In this section we describe protocol Syn-Phase-Spread. As mentioned above, for clarity's sake we assume that Syn-Phase-Spread uses 4-bit messages, and we show how to remove this assumption in Section 9.3.2. Three of such bits are devoted to the execution Syn-Clock, in order to synchronize a clock C modulo $2\lceil\gamma_{phase}\log n\rceil + \gamma_{phase}\lceil 2\log n\rceil$ for some constant γ_{phase}

[9]To get such upper bound, the analysis in [**FHK14**] leveraged on the property that in the $\mathcal{PUSH}$ model the number of agents getting a certain message can be upper bounded by the number of agent sending such message, which is not the case for the passive communication of the $\mathcal{PULL}$ model.

large enough. Throughout this section we assume, thanks to Theorem 16, that C has already been synchronized, which happens after $\tilde{\mathcal{O}}(\log n)$ rounds from the start of the protocol. In Section 9.3.1.1, we present a protocol Phase-Spread solving majority bit dissemination assuming agents already share a common clock.

9.3.1.1. *Protocol* Phase-Spread. Let γ_{phase} be a constant to be set later. Protocol Phase-Spread is executed periodically over periods of length $2\lceil\gamma_{phase}\log n\rceil + \gamma_{phase}\lceil 2\log n\rceil$, given by a clock C. One run of length $2\lceil\gamma_{phase}\log n\rceil + \gamma_{phase}\lceil 2\log n\rceil$ is divided in $2 + \lceil 2\log n\rceil$ phases, the first and the last ones lasting $\lceil\gamma_{phase}\log n\rceil$ rounds, all the other $\lceil 2\log n\rceil$ phases lasting γ_{phase} rounds. The first phase is called *boosting*, the last one is called *polling*, and all the intermediate ones are called *spreading*. For technical convenience, in Phase-Spread agents disregard the messages they get as their second pull[10].

During the boosting and the spreading phases, we make use of the parity of time to emulate the ability to actively send a message or to not-communicate anything as in the $\mathcal{PUSH}$ model[11]. In the first case we say that the agent is *speaking*, in the second case we say that the agent is *silent*. This induces a factor 2 slowdown which we henceforth omit for simplicity.

At the beginning of the boosting, each non-source agent u is silent. During the boosting and during each spreading phase, each silent agent pulls until she sees a speaking agent. When a silent agent u sees a speaking agent v, u memorizes $b_1(v)$ but remains silent until the end of the phase; at the end of the current phase, u starts speaking and sets $b_1(u) = b_1(v)$. The bit b_1 is then never modified until the clock C reaches 0 again. Then, during the polling phase, each agent u counts how many agents with $b_1 = 1$ and how many with $b_1 = 0$ she sees. At the end of the phase, each agent u sets their output bit to the most frequent value of b_1 observed during the polling phase. We want to show that, for all agents, the latter is b_{maj}, w.h.p. (i.e. the most frequent initial opinion among sources).

[10] In other words, Phase-Spread works in the $\mathcal{PULL}(1)$ model.

[11] Of course, agents are still not able to control who sees/contacts them.

PHASE-SPREAD protocol

1: If u is not speaking and the current phase is either the boosting or the spreading one, u does the following:
2: | u observes a random agent v.
3: | If v is speaking, u sets $b_1(u)$ equal to $b_1(v)$, and u will be speaking from the next phase.
4: | u sets c_0 and c_1 equal to 0.
5: If the current phase is polling:
6: | u observes a random agent v.
7: | If $b_1(v) = 1$, u increments c_1, otherwise increment c_0.
8: u outputs 1 if and only if $c_1 > c_0$.

Algorithm 6. The protocol PHASE-SPREAD, executed by each agent u.

9.3.1.2. *Analysis of* PHASE-SPREAD. We prove that at the end of the last spreading phase all agents are speaking and each agent has $b_1 = 1$ with probability $\frac{1}{2} + \varepsilon_{end}$ for some positive constant $\varepsilon_{end} = \varepsilon_{end}(\gamma_{phase}, \varepsilon)$ (where the dependency in γ_{phase} is monotonically increasing), $b_1 = 0$ otherwise, w.h.p. From the Chernoff bound (Corollary 9) and the union bound, this implies that when $\gamma_{phase} > 8/\varepsilon_{end}$ at the end of the polling phase each agent learns b_{maj}, w.h.p.

Without loss of generality, let $b_{maj} = 1$, i.e. $k_1 > k_0$. The analysis is divided in the following lemmas.

LEMMA 73. *At the end of the boosting phase it holds w.h.p.*

$$
\begin{aligned}
(168) \qquad & k_1^{(1)} + k_0^{(1)} \\
& \geqslant (k_1 + k_0)\,\frac{\gamma_{phase}}{3}\log n \cdot \mathbf{1}_{\left\{k_1+k_0 < \frac{n}{2\gamma_{phase}\log n}\right\}} \\
& \quad + \left(n\left(1 - \frac{1}{\sqrt{e}}\right) + \frac{1}{\sqrt{e}}\,(k_1 + k_0) - \sqrt{n\log n}\right) \\
& \qquad \cdot \mathbf{1}_{\left\{\frac{n}{2\gamma_{phase}} \leqslant k_1+k_0 \leqslant n - 2\sqrt{n\log n}\right\}} \\
& \quad + n\,\mathbf{1}_{\left\{k_1+k_0 > n-2\sqrt{n\log n}\right\}}, \\
(169) \qquad & \frac{k_1^{(1)}}{k_0^{(1)}} \geqslant \frac{k_1}{k_0}\left(1 - \sqrt{\frac{9}{\gamma_{phase}k_0}}\right).
\end{aligned}
$$

PROOF. First, we prove (168). By using the fact that if $|x| < 1$, it holds

$$
(170) \qquad e^{\frac{x}{1+x}} \leqslant 1 + x \leqslant e^x \leqslant 1 + \frac{x}{1-x}.
$$

we have

$$\begin{aligned}
&\mathbb{E}\left[k_1^{(1)} + k_0^{(1)}\right] \\
&= k_1 + k_0 + (n - k_1 - k_0)\left(1 - \left(1 - \frac{k_1 + k_0}{n}\right)^{\gamma_{phase}\log n}\right) \\
&\geqslant k_1 + k_0 + (n - k_1 - k_0)\left(1 - \exp\left(-\frac{k_1 + k_0}{n}\gamma_{phase}\log n\right)\right). \qquad (171)
\end{aligned}$$

We distinguish three cases.

Case $k_1 + k_0 < \frac{n}{2\gamma_{phase}\log n}$. By using (170) again, from (171) we get

$$\begin{aligned}
&\mathbb{E}\left[k_1^{(1)} + k_0^{(1)}\right] \\
&\geqslant k_1 + k_0 + (n - k_1 - k_0)\left(1 - \exp\left(-\frac{k_1 + k_0}{n}\gamma_{phase}\log n\right)\right) \\
&\geqslant k_1 + k_0 + (n - k_1 - k_0)\frac{\frac{k_1+k_0}{n}\gamma_{phase}\log n}{1 + \frac{k_1+k_0}{n}\gamma_{phase}\log n} \\
&\geqslant k_1 + k_0 + (n - k_1 - k_0)\frac{k_1 + k_0}{n}\frac{\gamma_{phase}}{2}\log n \\
&\geqslant k_1 + k_0 + \left(1 - \frac{k_1 + k_0}{2n}\right)(k_1 + k_0)\frac{\gamma_{phase}}{2}\log n \\
&\geqslant (k_1 + k_0)\left(1 + \left(1 - \frac{1}{4\gamma_{phase}\log n}\right)\frac{\gamma_{phase}}{2}\log n\right) \\
&\geqslant (k_1 + k_0)\frac{\gamma_{phase}}{2}\log n. \qquad (172)
\end{aligned}$$

From the Chernoff bound (Lemma 25), we thus get that w.h.p.

$$k_1^{(1)} + k_0^{(1)} \geqslant (k_1 + k_0)\frac{\gamma_{phase}}{3}\log n.$$

Case $\frac{n}{2\gamma_{phase}\log n} \leqslant k_1 + k_0 \leqslant n - 2\sqrt{n\log n}$. From (171), we have

$$\begin{aligned}
&\mathbb{E}\left[k_1^{(1)} + k_0^{(1)}\right] \\
&\geqslant k_1 + k_0 + (n - k_1 - k_0)\left(1 - \exp\left(-\frac{k_1 + k_0}{n}\gamma_{phase}\right)\right) \\
&\geqslant k_1 + k_0 + (n - k_1 - k_0)\left(1 - \frac{1}{\sqrt{e}}\right) \geqslant n\left(1 - \frac{1}{\sqrt{e}}\right) + \frac{k_1 + k_0}{\sqrt{e}}.
\end{aligned}$$

From the Chernoff bound (Lemma 25), we thus get that w.h.p.

$$k_1^{(1)} + k_0^{(1)} \geqslant n\left(1 - \frac{1}{\sqrt{e}}\right) + \frac{k_1 + k_0}{\sqrt{e}} - \sqrt{n\log n}.$$

Case $k_1^{(1)} + k_0^{(1)} > n - 2\sqrt{n \log n}$. The probability that a silent agent does not observe a speaking one is

$$\left(\frac{n - k_1 - k_0}{n}\right)^{\gamma_{phase} \log n} \leqslant \left(\frac{4 \log n}{n}\right)^{\frac{1}{2}\gamma_{phase} \log n},$$

hence by a simple union bound it follows that all agents are speaking, w.h.p.

Now, we prove (169). As before, we have two cases. The first case,

$$\frac{k_1}{k_0} \geqslant \frac{n}{2\gamma_{phase} \log n},$$

is a simple consequence of the Chernoff bound (Lemma 25).

In the second case,

$$\frac{k_1}{k_0} < \frac{n}{2\gamma_{phase} \log n},$$

let us consider the set of agents S_{boost} that start speaking at the end of the boosting, i.e. that observe a speaking agent during the phase. Observe that

$$|S_{boost}| = k_1^{(1)} - k_1 + k_0^{(1)} - k_0.$$

The probability that an agent in S_{boost} observes an agent in $\mathcal{B}$ (resp. $\mathcal{W}$) is $\frac{k_1}{k_1+k_0}$ (resp. $\frac{k_0}{k_1+k_0}$). Thus

$$\mathbb{E}\left[k_1^{(1)}\right] = k_1 + \frac{k_1}{k_1 + k_0}\mathbb{E}\left[|S_{boost}|\right] \quad \text{and}$$

$$\mathbb{E}\left[k_0^{(1)}\right] = k_0 + \frac{k_0}{k_1 + k_0}\mathbb{E}\left[|S_{boost}|\right]. \tag{173}$$

In particular

$$\frac{\mathbb{E}\left[k_1^{(1)}\right]}{\mathbb{E}\left[k_0^{(1)}\right]} = \frac{k_1 + \frac{k_1}{k_1+k_0}\mathbb{E}\left[|S_{boost}|\right]}{k_0 + \frac{k_0}{k_1+k_0}\mathbb{E}\left[|S_{boost}|\right]} = \frac{k_1}{k_0}, \tag{174}$$

and from (172) and (173) we have

$$\begin{aligned}
\mathbb{E}\left[k_0^{(1)}\right] &\geqslant \frac{k_0}{k_1 + k_0}\mathbb{E}\left[|S_{boost}|\right] \\
&= \frac{k_0}{k_1 + k_0}\left(\mathbb{E}\left[k_1^{(1)} + k_0^{(1)}\right] - (k_1 + k_0)\right) \\
&\geqslant (1 - o(1))\frac{k_0}{k_1 + k_0}\frac{\gamma_{phase}}{2}(k_1 + k_0)\log n \\
&= (1 - o(1))k_0\frac{\gamma_{phase}}{2}\log n, \qquad (175)
\end{aligned}$$

where the lower bound follows from the assumption $\frac{k_1}{k_0} < \frac{n}{2\gamma_{phase}\log n}$ and (172). From (175) and the multiplicative form of the Chernoff bound (Corollary 9), we have that w.h.p.

$$k_1^{(1)} \geqslant \mathbb{E}\left[k_1^{(1)}\right] - \sqrt{\mathbb{E}\left[k_1^{(1)}\right]\log n} \quad \text{and}$$

$$k_0^{(1)} \leqslant \mathbb{E}\left[k_0^{(1)}\right] + \sqrt{\mathbb{E}\left[k_0^{(1)}\right]\log n}. \tag{176}$$

Thus, since (173) implies $\mathbb{E}\left[k_1^{(1)}\right] \geqslant \mathbb{E}\left[k_0^{(1)}\right]$, we have

$$\begin{aligned}
\frac{k_1^{(1)}}{k_0^{(1)}} &\geqslant \frac{\mathbb{E}\left[k_1^{(1)}\right] - \sqrt{\mathbb{E}\left[k_1^{(1)}\right]\log n}}{\mathbb{E}\left[k_0^{(1)}\right] + \sqrt{\mathbb{E}\left[k_0^{(1)}\right]\log n}} \\
&= \frac{\mathbb{E}\left[k_1^{(1)}\right]}{\mathbb{E}\left[k_0^{(1)}\right]} \cdot \frac{1 - \sqrt{\frac{\log n}{\mathbb{E}\left[k_1^{(1)}\right]}}}{1 + \sqrt{\frac{\log n}{\mathbb{E}\left[k_0^{(1)}\right]}}} \\
&\geqslant \frac{\mathbb{E}\left[k_1^{(1)}\right]}{\mathbb{E}\left[k_0^{(1)}\right]} \cdot \left(1 - \sqrt{\frac{\log n}{\mathbb{E}\left[k_1^{(1)}\right]}} - \sqrt{\frac{\log n}{\mathbb{E}\left[k_0^{(1)}\right]}}\right) \\
&\geqslant \frac{\mathbb{E}\left[k_1^{(1)}\right]}{\mathbb{E}\left[k_0^{(1)}\right]} \cdot \left(1 - 2\sqrt{\frac{\log n}{\mathbb{E}\left[k_0^{(1)}\right]}}\right) \\
&= \frac{k_1}{k_0} \cdot \left(1 - \sqrt{\frac{9}{k_0\gamma_{phase}}}\right), \qquad (177)
\end{aligned}$$

concluding the proof. ∎

LEMMA 74. *At the end of the $i+1$th spreading phase, the following holds w.h.p.*

$$\begin{aligned}
&k_1^{(i+1)} + k_0^{(i+1)} \\
&\geqslant \left(k_1^{(i)} + k_0^{(i)}\right)\frac{\gamma_{phase}}{3}\mathbf{1}_{\left\{k_1^{(i)}+k_0^{(i)} < \frac{n}{2\gamma_{phase}}\right\}} \\
&\quad + \left(n\left(1 - \frac{1}{\sqrt{e}}\right) + \frac{1}{\sqrt{e}}\left(k_1^{(i)} + k_0^{(i)}\right) - \sqrt{n\log n}\right) \\
&\quad\quad \cdot \mathbf{1}_{\left\{\frac{n}{2\gamma_{phase}} \leqslant k_1^{(i)}+k_0^{(i)} \leqslant n - 2\sqrt{n\log n}\right\}} \\
&\quad + n\mathbf{1}_{\left\{k_1^{(i)}+k_0^{(i)} > n - 2\sqrt{n\log n}\right\}}, \qquad (178)
\end{aligned}$$

$$\frac{k_1^{(i+1)}}{k_0^{(i+1)}} \geqslant \frac{k_1^{(i)}}{k_0^{(i)}} \left(1 - 4\sqrt{\frac{\log n}{\gamma_{phase} k_0^{(i)}}}\right). \tag{179}$$

PROOF. The proof is almost the same as that of Lemma 73. From (170), we have

$$\begin{aligned}
&\mathbb{E}\left[k_1^{(i+1)} + k_0^{(i+1)}\right] \\
&= k_1^{(i)} + k_0^{(i)} + \left(n - k_1^{(i)} - k_0^{(i)}\right)\left(1 - \left(1 - \frac{k_1^{(i)} + k_0^{(i)}}{n}\right)^{\gamma_{phase}}\right) \\
&\geqslant k_1^{(i)} + k_0^{(i)} + \left(n - k_1^{(i)} - k_0^{(i)}\right)\left(1 - \exp\left(-\gamma_{phase}\frac{k_1^{(i)} + k_0^{(i)}}{n}\right)\right). \qquad (180)
\end{aligned}$$

We distinguish three cases.

Case $k_1^{(i)} + k_0^{(i)} < \frac{n}{2\gamma_{phase}}$. From (170) and (180) we get

$$\begin{aligned}
&\mathbb{E}\left[k_1^{(i+1)} + k_0^{(i+1)}\right] \\
&\geqslant k_1^{(i)} + k_0^{(i)} + \left(n - k_1^{(i)} - k_0^{(i)}\right) \\
&\quad \cdot \left(1 - \exp\left(-\frac{k_1^{(i)} + k_0^{(i)}}{n}\gamma_{phase}\right)\right) \\
&\geqslant k_1^{(i)} + k_0^{(i)} + \left(n - k_1^{(i)} - k_0^{(i)}\right) \cdot \frac{\frac{k_1^{(i)}+k_0^{(i)}}{n}\gamma_{phase}}{1 + \frac{k_1^{(i)}+k_0^{(i)}}{n}\gamma_{phase}} \\
&\geqslant k_1^{(i)} + k_0^{(i)} + \left(n - k_1^{(i)} - k_0^{(i)}\right) \cdot \frac{k_1^{(i)} + k_0^{(i)}}{2n}\gamma_{phase} \\
&\geqslant k_1^{(i)} + k_0^{(i)} + \left(1 - \frac{k_1^{(i)} + k_0^{(i)}}{n}\right) \cdot \left(k_1^{(i)} + k_0^{(i)}\right)\frac{\gamma_{phase}}{2} \\
&\geqslant \left(k_1^{(i)} + k_0^{(i)}\right)\left(1 + \left(1 - \frac{1}{2\gamma_{phase}}\right)\frac{\gamma_{phase}}{2}\right) \\
&\geqslant \left(k_1^{(i)} + k_0^{(i)}\right)\frac{\gamma_{phase}}{2}. \qquad (181)
\end{aligned}$$

After the boosting phase, i.e. for $i \geqslant 1$, it follows from Lemma 73 that $k_1^{(i)} + k_0^{(i)} = \Omega\left(\gamma_{phase} \log n\right)$. From the Chernoff bound (Lemma 25), if γ_{phase} is chosen big enough, we thus get that w.h.p.

$$k_1^{(i+1)} + k_0^{(i+1)} \geqslant \left(k_1^{(i)} + k_0^{(i)}\right)\frac{\gamma_{phase}}{3}.$$

Case $\frac{n}{2\gamma_{phase}} \leqslant k_1^{(i)} + k_0^{(i)} \leqslant n - 2\sqrt{n \log n}$. From (180), we have

$$\begin{aligned}
&\mathbb{E}\left[\left(k_1^{(i+1)} + k_0^{(i+1)}\right)\right] \\
&\geqslant k_1^{(i)} + k_0^{(i)} + \left(n - k_1^{(i)} - k_0^{(i)}\right) \\
&\qquad \cdot \left(1 - \exp\left(-\frac{k_1^{(i)} + k_0^{(i)}}{n}\gamma_{phase}\right)\right) \\
&\geqslant k_1^{(i)} + k_0^{(i)} + \left(n - k_1^{(i)} - k_0^{(i)}\right)\left(1 - \frac{1}{\sqrt{e}}\right) \\
&\geqslant n\left(1 - \frac{1}{\sqrt{e}}\right) + \frac{1}{\sqrt{e}}\left(k_1^{(i)} + k_0^{(i)}\right).
\end{aligned}$$

From the Chernoff bound (Lemma 25), we thus get that w.h.p.

$$k_1^{(i+1)} + k_0^{(i+1)} \geqslant n\left(1 - \frac{1}{\sqrt{e}}\right) + \frac{1}{\sqrt{e}}\left(k_1^{(i)} + k_0^{(i)}\right) - \sqrt{n \log n}.$$

Case $k_1^{(i)} + k_0^{(i)} > n - 2\sqrt{n \log n}$. The probability that a silent agent does not observe a speaking one is

$$\left(\frac{n - k_1^{(i)} - k_0^{(i)}}{n}\right)^{\gamma_{phase}} \leqslant \left(\frac{4 \log n}{n}\right)^{\frac{1}{2}\gamma_{phase}},$$

hence by a simple union bound it follows that all agents are speaking, w.h.p.

Now, we prove (179). As in the proof of (169), we have two cases. The first case, $\frac{k_1}{k_0} \geqslant \frac{n}{2\gamma_{phase}}$, is a simple consequence of the Chernoff bound (Lemma 25). Otherwise, let us assume $\frac{k_1}{k_0} < \frac{n}{2\gamma_{phase}}$. With an analogous argument to that for (173) and (174) we can prove

$$\frac{\mathbb{E}\left[k_1^{(i+1)}\right]}{\mathbb{E}\left[k_0^{(i+1)}\right]} = \frac{k_1^{(i)}}{k_0^{(i)}}, \tag{182}$$

and

$$\begin{aligned}
\mathbb{E}\left[k_1^{(i+1)}\right] &= k_1^{(i)} + \frac{k_1^{(i)}}{k_1^{(i)} + k_1^{(i)}}\mathbb{E}\left[k_1^{(i+1)} - k_1^{(i)} + k_0^{(i+1)} - k_0^{(i)}\right], \\
\mathbb{E}\left[k_0^{(i+1)}\right] &= k_0^{(i)} + \frac{k_0^{(i)}}{k_1^{(i)} + k_0^{(i)}}\mathbb{E}\left[k_1^{(i+1)} - k_1^{(i)} + k_0^{(i+1)} - k_0^{(i)}\right]. \qquad (183)
\end{aligned}$$

As in (176), from the multiplicative form of the Chernoff bound (Corollary 9) we have that w.h.p.

$$k_1^{(i+1)} \geqslant \mathbb{E}\left[k_1^{(i+1)}\right] - \sqrt{\mathbb{E}\left[k_1^{(i+1)}\right]\log n} \quad \text{and}$$

$$k_0^{(i+1)} \leqslant \mathbb{E}\left[k_0^{(i+1)}\right] + \sqrt{\mathbb{E}\left[k_0^{(i+1)}\right]\log n}. \tag{184}$$

Thus, as in (177), from (184) and (182), we get

$$\begin{aligned}\frac{k_1^{(i+1)}}{k_0^{(i+1)}} &\geqslant \frac{\mathbb{E}\left[k_1^{(i+1)}\right]}{\mathbb{E}\left[k_0^{(i+1)}\right]} \cdot \left(1 - 2\sqrt{\frac{\log n}{\mathbb{E}\left[k_0^{(i+1)}\right]}}\right) \\ &\geqslant \frac{k_1^{(i)}}{k_0^{(i)}} \cdot \left(1 - 4\sqrt{\frac{\log n}{\gamma_{phase} k_0^{(i)}}}\right),\end{aligned}$$

where, as in (175), in the last inequality we used that from (181) and (183) it holds

$$\mathbb{E}\left[k_0^{(i+1)}\right] \geqslant \frac{\gamma_{phase}}{4} k_0^{(i)}.$$

gg

From the previous two lemmas, we can derive the following proposition, which concludes the proof.

Proposition 3. *If $k_1 \geqslant k_0(1+\varepsilon)$ for some constant $\varepsilon > 0$, then at the end of the last spreading phase it holds w.h.p.*

$$k_1^{(1+2\log n)} = n - k_0^{(1+2\log n)} \geqslant k_0^{(1+2\log n)}\left(1+\varepsilon_{end}\right), \tag{185}$$

where $\varepsilon_{end} = \varepsilon - \frac{4}{\sqrt{\gamma_{phase}}}$.

Proof. We first show how the equality in (185) follows from (178). When

$$k_1^{(i)} + k_0^{(i)} < \frac{n}{2\gamma_{phase}},$$

(178) shows that $k_1^{(i)} + k_0^{(i)}$ increases by multiplicative a factor γ_{phase} at the end of each spreading phase. When

$$\frac{n}{2\gamma_{phase}} \leqslant k_1^{(i)} + k_0^{(i)} \leqslant n - 2\sqrt{n\log n},$$

(178) shows that

$$n - k_1^{(i+1)} - k_0^{(i+1)} \leqslant \frac{n - k_1^{(i)} - k_0^{(i)}}{\sqrt{e}} - \sqrt{n\log n} \leqslant \frac{n - k_1^{(i)} - k_0^{(i)}}{\sqrt{e}}.$$

Hence the number of silent agents decreases by a factor $\sqrt{e}$ after each spreading phase. Lastly, when

$$k_1^{(i)} + k_0^{(i)} > n - 2\sqrt{n\log n},$$

after one more spreading phase, a simple application of the union bound shows that $k_1^{(i+1)} + k_0^{(i+1)}$ is equal to n, w.h.p. As a consequence, if γ_{phase} is big enough, after less than $1 + 2\log n$ spreading phases it holds that w.h.p.

$$k_1^{(1+2\log n)} = n - k_0^{(1+2\log n)}.$$

The inequality in (185) can be derived from (179), as follows. From (169) and (179) we have

$$\frac{k_1^{(1+2\log n)}}{k_0^{(1+2\log n)}} \geqslant \frac{k_1}{k_0}\left(1 - \sqrt{\frac{9}{\gamma_{phase}k_0}}\right)\prod_{i=2}^{1+2\log n}\left(1 - \sqrt{\frac{16\log n}{\gamma_{phase}k_0^{(i)}}}\right). \tag{186}$$

We can estimate the product as

$$\begin{aligned}
&\prod_{i=2}^{1+2\log n}\left(1 - \sqrt{\frac{16\log n}{\gamma_{phase}k_0^{(i)}}}\right)\\
&\geqslant \exp\left(-4\sum_{i=2}^{1+2\log n}\frac{1}{\left(\sqrt{\gamma_{phase}}\right)^i}\right),\\
&\geqslant \exp\left\{4\left(1 + \frac{1}{\sqrt{\gamma_{phase}}} - \frac{1 - (\gamma_{phase})^{-\frac{2+2\log n}{2}}}{1 - (\gamma_{phase})^{-\frac{1}{2}}}\right)\right\}\\
&\geqslant \exp\left\{-4\left(\frac{1}{\gamma_{phase} - \sqrt{\gamma_{phase}}} - n^{-\frac{2\log\gamma_{phase}}{2}}\right)\right\}\\
&\geqslant \left(1 - \frac{5}{\gamma_{phase}}\right), \qquad (187)
\end{aligned}$$

where in the first and last inequality we used that $1 - x \geqslant e^{-\frac{x}{1-x}}$ if $|x| < 1$.

Finally, from (186) and (187) we get

$$\begin{aligned}
\frac{k_1^{(1+2\log n)}}{k_0^{(1+2\log n)}} &\geqslant \frac{k_1}{k_0}\left(1 - \sqrt{\frac{9}{\gamma_{phase}k_0}}\right)\left(1 - \frac{5}{\gamma_{phase}}\right)\\
&\geqslant \frac{k_1}{k_0}\left(1 - \frac{4}{\sqrt{\gamma_{phase}}}\right),
\end{aligned}$$

concluding the proof. ∎

Having completed the proof of Proposition 3, in the next Section we can finally prove the main theorem of this chapter.

9.3.2. Proof of Theorem 15

THEOREM 15 (SYN-PHASE-SPREAD). *Fix an arbitrarily small constant $\varepsilon > 0$. There exists a protocol, called* SYN-PHASE-SPREAD*, which solves the majority bit dissemination problem in a self-stabilizing manner in $\tilde{\mathcal{O}}(\log n)$*

rounds[12]*, w.h.p. using* 3*-bit messages, provided that the majority bit is supported by at least a fraction* $\frac{1}{2}+\varepsilon$ *of the source agents.*

PROOF OF THEOREM 15. From Proposition 3, it follows that at the end of the last spreading phase, all agents have been informed. After the last spreading phase, during the polling phase, each agent samples $\gamma_{phase}\log n$ opinions from the population and then adopts the majority of these as her output bit. Thus, (185) ensures that each sample holds the correct opinion with probability $\geqslant \frac{1}{2}+\varepsilon_{end}$. Hence, by the Chernoff bound (Lemma 76) and a union bound, if γ_{phase} is big enough then the majority of the $\gamma_{phase}\log n$ samples corresponds to the correct value for all the n agents, w.h.p.

The protocol obtained so far solves majority bit dissemination, but it does it using 4 bits per message rather than 3. Indeed, synchronizing a clock using SYN-CLOCK takes 3 bits, and we use an extra bit to execute PHASE-SPREAD described in Section 9.3.1.1. However, the protocol SYN-PHASE-SPREAD has the independence property. This follows from Lemma 70 with SYN-GENERIC =SYN-CLOCK, P =PHASE-SPREAD, SYN-P =SYN-PHASE-SPREAD, together with the observation that PHASE-SPREAD is self-stabilizing. We can thus reduce the message length of SYN-PHASE-SPREAD to 3 bits using again the Message Reduction Theorem, with a time overhead of a factor 4 only. ▣

9.4. Proofs of Technical Lemmas

LEMMA 75. *Let* $f,g:\mathbb{R}_+\to\mathbb{R}$ *be functions defined by* $f(x)=\lceil\log x\rceil+1$ *and*

$$\tau(x)=\inf\left\{k\in\mathbb{N}\mid f^{\circledast k}(x)\leqslant 3\right\},$$

where we denote by $f^{\circledast k}$ *the* k*-fold iteration of* f*. It holds that*

$$\tau(T)\leqslant \log^{\circledast 4} T+\mathcal{O}(1).$$

PROOF. We can notice that

$$f(T)\leqslant T-1,$$

if T is bigger than some constant c. Moreover, when $f(x)\leqslant c$, the number of iterations before reaching 1 is $O(1)$. This implies that

$$\tau(T)\leqslant T+\mathcal{O}(1).$$

But in fact, by definition, $\ell(T)=g\left(f^{\circledast 4}(T)\right)+4$ (provided $f^{\circledast 4}(T)>1$, which holds if T is big enough). Hence

$$\tau(T)\leqslant g\left(f^{\circledast 4}(T)\right)+4\leqslant f^{\circledast 4}(T)+\mathcal{O}(1)\leqslant \log^{\circledast 4} T+\mathcal{O}(1).$$

▣

[12]With a slight abuse of notation, with $\tilde{\mathcal{O}}(f(n)g(T))$ we refer to $f(n)g(T)\cdot\log^{\mathcal{O}(1)}(f(n))\cdot\log^{\mathcal{O}(1)}(g(T))$. All logarithms are in base 2.

CHAPTER 10

Open Problems

As discussed in Chapter 1, the Averaging dynamics studied in Chapter 4 has the disadvantages of assuming that agents can interpret their state as a real number and perform arithmetical operations. Furthermore, the dynamics operates in the $\mathcal{LOCAL}$ model [**Pel00**].

It is an important open problem whether the Averaging dynamics itself can still achieve the same performances when implemented in the important context of population protocols: rather than having all nodes computing the average of all neighbors at each round, at each time step only a single edge is sampled u.a.r., and its two endpoints averages their values.

An even greater open problem is whether the community detection problem can be solved, with comparable performances, by even simpler dynamics, such as the 3-Majority dynamics, which relies only on the ability to compute the mode of a sample, i.e. on testing equality.

Regarding the upper bound on the convergence time of the 3-Majority dynamics given in Chapter 5 (Theorem 8), we believe that it is not tight w.r.t. k. We think that at least a factor k can be saved. To this aim, we would need to show that "more" opinions get small during a phase. This number should also depend on the current number of big colors. Another idea would be that of (also) considering the growth of the maximal opinion. Unfortunately, differently from the minimal opinion (see (74) in Section 5.4), we have no good bound on the expected drift for the maximal opinion that holds from *any* configuration. So, we don't see how to efficiently adapt our approach without this crucial ingredient.

A more general open question is to analyze stabilizing almost-consensus dynamics, such as the 3-Majority one, in some interesting graph topologies. We believe that a suitable combination of our analysis and some previous analysis for the binary case [**CER14**] might be useful on expander graphs [**HLW06**] and some classes of random evolving graphs [**CMM$^+$10**].

Moving to Chapter 6, we believe that the monochromatic distance investigated there might represent a general lower bound on the convergence time of *any* plurality dynamics which uses only $\log k + \Theta(1)$ bits of local memory.

Analogously to the 3-Majority dynamics, an interesting future research is the study of the Undecided-State dynamics (or other simple dynamics) for solving the plurality consensus problem over other classes of graphs in $o(k)$ time. In this work, we combined this dynamics with parallel random walks in order to get an efficient protocol for regular expander graphs. We

believe that similar protocols can work also in other classes of graphs such as Erdős-Rényi graphs and dynamic graphs [**CMM+10, CCD+13**].

In Chapter 7, we showed that the repeated balls-into-bin process, which models parallel random walks in the $\mathcal{PUSH}$ model on a complete graph, is self-stabilizing when the number m of balls equals the number n of bins (obviously, this is still the case, whenever $m < n$). An interesting open question is whether this result extends to larger values of m, i.e., for any $m = \mathcal{O}(n \log n)$. We believe an approach based on a lower bound on the number of empty bins might still work. Simulation results for increasing values of n (up to $n \sim 10^5$) show that the number of empty bins is still compatible with a linear function, even if standard deviation in our experiments turned out to be relatively large.

A more general interesting question is the study of parallel random walks in the uniform $\mathcal{PUSH}$ model over more general graph classes. This line of research is also motivated by several applications of the process [**BCEG10, Coo11, EK15, HPP+16**]. The analysis of this process in Section 6.3 provides a bound $\mathcal{O}(\sqrt{t})$ on the maximum load after t rounds on regular graphs [**BCN+15a**]. We believe this previous bound for regular graphs is far from tight and it leads to rough bounds on parallel cover times on these networks. We conjecture that the maximum load remains logarithmic for a long period in any *regular* graph. A possible reason for this phenomenon (if true) might be that the expected difference between (token) arrivals and departures is always *non-positive* at every node in regular graphs. As highlighted in our analysis of the complete graph, this fact alone is not enough but it could be combined with a suitable bound on the number of empty bins, in order to prove our conjecture in this more general case. Unfortunately, non-complete graphs present a further technical issue: in order to apply any argument based on the presence of empty bins, not only do we need to argue about their number, but also about their distribution across the network. This technical issue seems to be far from trivial even on simple topologies such as rings.

Another interesting question concerning the repeated balls-into-bins process is the tightness of the bound on the maximum load provided by Theorem 12. In the classical (one shot) balls-into-bins problem, it is well-known that the maximum load of the bins is $\Theta(\log n / \log \log n)$, w.h.p. One may wonder whether our $\mathcal{O}(\log n)$ upper bound on the maximum load of the repeated process for a polynomial number of rounds is tight, or it can be improved to $\mathcal{O}(\log n / \log \log n)$. We conjecture that, within any polynomial time window, the probability that the maximum load asymptotically exceeds $\log n / \log \log n$ is non-negligible.

In Chapter 8, we solved the general version of bit dissemination and plurality consensus in biological systems. That is, we have solved these problems for an arbitrarily large number k of opinions. We are not aware of realistic biological contexts in which the number of opinions might be a function of the number n of individuals. Nevertheless, it could be interesting, at least

from a conceptual point to view, to address bit dissemination and plurality consensus in a scenario in which the number of opinions varies with n. This appears to be a technically challenging problem. Indeed, extending the results in the extended abstract of [**FHK15**] from 2 opinions to any constant number k of opinions already required to use complex tools. Yet, several of these tools do not apply if k depends on n. This is typically the case of Proposition 1. We let as an open problem the design of stochastic tools enabling to handle the scenario where $k = k(n)$.

In Chapter 9 we have dealt with the construction of protocols in highly congested stochastic interaction patterns. Corresponding challenges are particularly evident when it is difficult to guarantee synchronization, which seems to be essential for emulating a typical protocol that relies on many bits per message with a protocol that uses fewer bits. Chapter 9 shows that in the $\mathcal{PULL}$ model, if a self-stabilizing protocol satisfies the bitwise-independence property then it can be emulated with only 3 bits per message. Using this rather general transformer, we solve the self-stabilizing clock synchronization and majority bit dissemination problems in almost-logarithmic time and using only 3 bits per message. It remains an open problem whether the message size of either one of these problems can be further reduced while keeping the running time polylogarithmic.

In particular, even for the self-stabilizing bit dissemination problem (with a single source) it remains open whether there exists a polylogarithmic protocol that uses a single bit per interaction. In fact, we investigated several candidate protocols which seem promising in experimental simulation, but appear to be out of reach of current techniques for analysing randomly-interacting agent systems in a self-stabilizing context. Let us informally present one of them, called BFS[1]. Let ℓ,$k \in \mathbb{N}$ be two parameters, say of order $O(\log n)$. Agents can be in 3 states: *boosting*, *frozen* or *sensitive*.

- Boosting agents behave as in the 3-Median dynamics protocol: they apply the majority rule to the 2 values they see in a given round and make it into their opinion for the next round. They also keep a counter T. If they have seen only agents of a given color b for ℓ rounds, they become sensitive to the opposite value.
- b-sensitive agents turn into frozen-b agents if they see value b.
- b-frozen agents keep the value b for k rounds before becoming boosters again.

Intuitively what we expect is that, from every configuration, at some point almost all agents would be in the boosting state. Then, the boosting behavior would lead the agents to converge to a value b (which depends on the initial conditions). Most agents would then become sensitive to $1-b$. If the source has opinion $1-b$ then there should be a switch from b to $1-b$. The "frozen" period is meant to allow for some delay in the times at which agents become sensitive, and then flip their opinion. This algorithm however, as the

[1]A similar protocol was suggested during discussions with Bernhard Haeupler.

other candidate protocols we have considered, does not seem amenable to a rigorous analysis in the $\mathcal{PULL}$ model which accounts for the self-stabilizing requirement.

APPENDIX A

Mathematical Tools

In this appendix we review several variants of the Chernoff bound used throughout this work. We also state a *reverse* version of the Chernoff bound and other technical tools we make use of in the proofs.

LEMMA 76 (Chernoff bounds). *Let $X = \sum_{i=1}^{n} X_i$ where X_i's are independent Bernoulli random variables and let*

$$\mu_L \leqslant \mu = \mathbb{E}[X] \leqslant \mu_H.$$

Then,

- *For any $0 < \delta \leqslant 4$,*

$$\Pr(X > (1+\delta)\mu) < e^{-\frac{\delta^2\mu}{4}}; \tag{188}$$

- *For any $\delta \geqslant 4$,*

$$\Pr(X > (1+\delta)\mu) < e^{-\delta\mu}; \tag{189}$$

- *For any $\lambda_U > 0$ and $\lambda_L \in (0,1)$,*

$$\Pr(X \geqslant \mu + \lambda) \leqslant e^{-2\frac{\lambda^2}{n}}, \tag{190}$$

$$\Pr(X \leqslant \mu - \lambda) \leqslant e^{-2\frac{\lambda^2}{n}}.$$

- *For any $\delta \in (0,1)$,*

$$\Pr(X \leqslant (1-\delta)\mu_L) \leqslant \exp\left(-\frac{\delta^2}{2}\mu_L\right), \tag{191}$$

$$\Pr(X \geqslant (1+\delta)\mu_H) \leqslant \exp\left(-\frac{\delta^2}{3}\mu_H\right). \tag{192}$$

THEOREM 25 ([**McD98**]). *Let $X_1, ..., X_n$ be n independent random variables. If $X_i \leqslant M$ for each i, then*

$$\Pr\left(\sum_i X_i \geqslant \mathbb{E}\left[\sum_i X_i\right] + \lambda\right) \leqslant \exp\left\{-\frac{\lambda^2}{2\left(\sqrt{\sum_i \mathbb{E}\left[X_i^2\right]} + \frac{M\lambda}{3}\right)}\right\}, \tag{193}$$

and if $X_i \geqslant -M$ for each i, then

(194)
$$\Pr\left(\sum_i X_i \leqslant \mathbb{E}\left[\sum_i X_i\right] - \lambda\right) \leqslant \exp\left\{-\frac{\lambda^2}{2\left(\sqrt{\sum_i \mathbb{E}\left[X_i^2\right]} + \frac{M\lambda}{3}\right)}\right\}.$$

COROLLARY 9. *If the X_is are binary, then for $\lambda = \sqrt{\mathbb{E}\left[\sum_i X_i\right] \log n}$,* (193) *and* (194) *become*

$$\Pr\left(\sum_i X_i \geqslant \mathbb{E}\left[\sum_i X_i\right] + \sqrt{\mathbb{E}\left[\sum_i X_i\right] \log n}\right) \leqslant e^{-\sqrt{\mathbb{E}\left[\sum_i X_i\right] \log n}},$$
$$\Pr\left(\sum_i X_i \leqslant \mathbb{E}\left[\sum_i X_i\right] - \sqrt{\mathbb{E}\left[\sum_i X_i\right] \log n}\right) \leqslant e^{-\sqrt{\mathbb{E}\left[\sum_i X_i\right] \log n}}.$$

PROOF. If the X_is are binary then $\sum_i \mathbb{E}\left[X_i^2\right] \leqslant \sum_i \mathbb{E}\left[X_i\right]$. Together with $\lambda = \sqrt{\mathbb{E}\left[\sum_i X_i\right] \log n}$, the latter fact implies that in the exponent of the right hand side of (193) and (194) we get

$$\begin{aligned}
&-\frac{\lambda^2}{2\left(\sqrt{\sum_i \mathbb{E}\left[X_i^2\right]} + \frac{M\lambda}{3}\right)} \\
&\leqslant -\frac{\mathbb{E}\left[\sum_i X_i\right] \log n}{2\left(\sqrt{\sum_i \mathbb{E}\left[X_i\right]} + \frac{1}{3}\sqrt{\mathbb{E}\left[\sum_i X_i\right] \log n}\right)} \\
&\leqslant -\frac{\mathbb{E}\left[\sum_i X_i\right] \log n}{2\sqrt{\mathbb{E}\left[\sum_i X_i\right]}\left(\frac{1}{3}\sqrt{\log n} + 1\right)} \leqslant -\sqrt{\mathbb{E}\left[\sum_i X_i\right] \log n}.
\end{aligned}$$

∎

LEMMA 77. *Let $\{X_t\}_{t\in[n]}$ be n i.i.d. random variables such that*

$$X_t = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } r, \\ -1 & \text{with probability } q. \end{cases}$$

with $p + r + q = 1$. It holds

$$\Pr\left(\sum_i X_t \leqslant (1-\theta)\,\mathbb{E}\left[\sum_i X_t\right] - \theta n\right) \leqslant \exp\left(-\frac{\theta^2}{4}\left(\mathbb{E}\left[\sum_i X_t\right] + n\right)\right).$$

PROOF. Let us define the r.v.

$$Y_t = \frac{X_t + 1}{2}. \tag{195}$$

We can apply the Chernoff-Hoeffding bound to Y_t (see Theorem 1.1 in [**DP09**]), obtaining

$$\Pr\left(\sum_i Y_i \leqslant (1-\theta)\,\mathbb{E}\left[\sum_i Y_i\right]\right) \leqslant \exp\left(-\frac{\theta^2}{2}\mathbb{E}\left[\sum_i Y_i\right]\right)$$

for any $\theta \in (0,1)$. Substituting (195) we have

$$\begin{aligned}\Pr\left(\sum_i X_t + n \leqslant (1-\theta)\left(\mathbb{E}\left[\sum_i X_t\right] + n\right)\right) \\ = \Pr\left(\sum_i X_t \leqslant (1-\theta)\,\mathbb{E}\left[\sum_i X_t\right] - \theta n\right) \\ \leqslant \exp\left(-\frac{\theta^2}{4}\left(\mathbb{E}\left[\sum_i X_t\right] + n\right)\right),\end{aligned}$$

concluding the proof. ∎

The following folklore "reverse"-Chernoff bound [**Mou14**, Theorem 2] shows that the Chernoff bound is essentially tight[1]

THEOREM 26 (Reverse Chernoff bound). *Let X be the sum of m independent Bernoulli variables with probability $p \leqslant 1/4$ and let $\mu = pm$. Then, for any $t \in (0, m-\mu)$:*

$$\Pr\left(X - \mu > t\right) \geqslant \frac{1}{4}e^{-\frac{2t^2}{\mu}}.$$

We are often interested in the expected value of a stochastic process at a time which is itself a random variable. Doob's Optional Stopping Theorem allows us to know such expected value, under suitable hypothesis.

THEOREM 27 ([**Doo53**], see also Corollary 17.8 in [**LPW09**] or Theorem 10.10 in [**Wil91**]). *Let $\{X_t\}_{t\in\mathbb{N}}$ be a discrete-time martingale and τ be a stopping time with values in $\mathbb{N}\cup\infty$, with respect to a given filtration $\mathcal{F}_t$. Assume one of the following three conditions holds:*

1. *The stopping time is almost surely bounded, i.e. $\Pr(\tau \leqslant c) = 1$ for some constant c;*
2. *$\mathbb{E}[\tau] < \infty$ and for some constant c*

$$\Pr(\mathbb{E}[|X_{t+1} - X_t| \,|\, \mathcal{F}_t] \leqslant c \,|\, \{\tau > t\}) = 1,$$

   for all $t \in \mathbb{N}$;
3. *For some constant c, $\Pr(|X_{\min\{\tau,t\}}| \leqslant c) = 1$ for all $t \in \mathbb{N}$.*

[1] A number of pretty similar results can be found in specialized mathematical forums, for example `http://cstheory.stackexchange.com/questions/14471/reverse-chernoff-bound`.

Then

$$\mathbb{E}[X_\tau] = \mathbb{E}[X_0].$$

REMARK 10. The proof of Theorem 27 naturally extends to supermartingales and submartingales.

Finally, the following fact is useful when dealing with many events in the uniform $\mathcal{PUSH}$ and $\mathcal{PULL}$ models.

FACT 1. *If* $f(n) = \omega(1)$ *and* $g(n) = o(f(n))$ *then*

$$\left(1 \pm \frac{1}{f(n)}\right)^{g(n)} = 1 \pm O\left(\frac{g(n)}{f(n)}\right).$$

Bibliography

[AAB+11] Y. Afek, N. Alon, O. Barad, E. Hornstein, N. Barkai, and Z. Bar-Joseph. A biological solution to a fundamental distributed computing problem. *Science*, 331(6014):183–185, 2011.

[AAD+06] D. Angluin, J. Aspnes, Z. Diamadi, M. J. Fischer, and R. Peralta. Computation in networks of passively mobile finite–state sensors. *Distributed computing*, 18(4):235–253, 2006.

[AAE06] D. Angluin, J. Aspnes, and D. Eisenstat. Stably computable predicates are semilinear. In *In Proceedings of the 25th Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC'06)*, pages 292–299. ACM, 2006.

[AAE08] D. Angluin, J. Aspnes, and D. Eisenstat. A simple population protocol for fast robust approximate majority. *Distributed Computing*, 21(2):87–102, 2008. (Preliminary version in DISC'07).

[AAFJ08] D. Angluin, J. Aspnes, M. J. Fischer, and H. Jiang. Self-stabilizing population protocols. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 3(4):13, 2008.

[AAK+08] N. Alon, C. Avin, M. Koucky, G. Kozma, Z. Lotker, and M. R. Tuttle. Many random walks are faster than one. In *Proceedings of the Twentieth Annual Symposium on Parallelism in Algorithms and Architectures (SPAA'08)*, pages 119–128. ACM, 2008.

[ABH14] E. Abbe, A. S Bandeira, and G. Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, 2014.

[ABKU99] Y. Azar, A. Z. Broder, A. R. Karlin, and E. Upfal. Balanced allocations. *SIAM journal on computing*, 29(1):180–200, 1999.

[ACMR95] M. Adler, S. Chakrabarti, M. Mitzenmacher, and L. Rasmussen. Parallel randomized load balancing. In *Proceedings of the 27th Annual ACM Symposium on Theory of Computing*, STOC'95, pages 238–247. ACM, 1995.

[AD15] M. A. Abdullah and M. Draief. Global majority consensus by local majority polling on graphs of a given degree sequence. *Discrete Applied Mathematics*, 180:1–10, 2015.

[AFJ06] D. Angluin, M. J. Fischer, and H. Jiang. Stabilizing consensus in mobile networks. In *Proceedings of Distributed Computing in Sensor Systems (DCOSS'06)*, volume 4026 of *LNCS*, pages 37–50, 2006.

[AG15] D. Alistarh and R. Gelashvili. Polylogarithmic-time leader election in population protocols. In *Proceedings, Part II, of the 42nd International Colloquium on Automata, Languages, and Programming*, volume 9135 of *ICALP'15*, pages 479–491. Springer-Verlag New York, Inc., 2015.

[AGV15] D. Alistarh, R. Gelashvili, and M. Vojnovic. Fast and exact majority in population protocols. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing*, PODC'15, pages 47–56. ACM, 2015.

[AHR96] H. Attiya, A. Herzberg, and S. Rajsbaum. Optimal clock synchronization under different delay assumptions. *SIAM Journal on Computing*, 25(2):369–389, 1996.

[AKL08] C. Avin, M. Koucký, and Z. Lotker. How to explore a fast-changing world (cover time of a simple random walk on evolving graphs). In *Proceedings of the 35th International Colloquium on Automata, Languages and Programming, Part I*, ICALP '08, pages 121–132. Springer-Verlag, 2008.

[AKU05] A. Anagnostopoulos, A. Kirsch, and E. Upfal. Load balancing in arbitrary network topologies with stochastic adversarial input. *SIAM Journal on Computing*, 34(3):616–639, 2005.

[AR07] J. Aspnes and E. Ruppert. An introduction to population protocols. *Bulletin of the EATCS*, 93:98–117, 2007.

[AS09] B. Awerbuch and C. Scheideler. Towards a scalable and robust dht. *Theory of Computing Systems*, 45(2):234–260, 2009.

[AS15] E. Abbe and C. Sandon. Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic bp, and the information-computation gap. *arXiv preprint arXiv:1512.09080*, 2015.

[Asm03] S. Asmussen. *Applied probability and queues*. Springer, 2003.

[Asp12] J. Aspnes. Faster randomized consensus with an oblivious adversary. In *Proceedings of the 31st Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC'12)*, pages 1–8. ACM, 2012.

[AYSS09] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione. Broadcast gossip algorithms for consensus. *IEEE Transactions on Signal Processing*, 57(7):2748–2761, 2009.

[BBK11] J. Beauquier, J. Burman, and S. Kutten. A self-stabilizing transformer for population protocols with covering. *Theoretical Computer Science*, 412(33):4247–4259, 2011.

[BC09] M. J. Barber and J. W. Clark. Detecting network communities by propagating labels under constraints. *Physical Review E*, 80(2):026129, 2009.

[BCEG10] P. Berenbrink, J. Czyzowicz, R. Elsässer, and L. Gasieniec. Efficient information exchange in the random phone-call model. In *Proceedings of the 37th International Colloquium on Automata, Languages, and Programming (ICALP'10)*, pages 127–138. Springer, 2010.

[BCN+14] L. Becchetti, A. Clementi, E. Natale, F. Pasquale, R. Silvestri, and L. Trevisan. Simple dynamics for plurality consensus. In *Proceedings of the 26th ACM Symposium on Parallelism in Algorithms and Architectures (SPAA'14)*, pages 247–256. ACM, 2014.

[BCN+15a] L. Becchetti, A. Clementi, E. Natale, F. Pasquale, and R. Silvestri. Plurality consensus in the gossip model. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'15)*, pages 371–390. SIAM, 2015.

[BCN+15b] L. Becchetti, A. Clementi, E. Natale, F. Pasquale, and L. Trevisan. Find your place: Simple distributed algorithms for community detection. *arXiv preprint arXiv:1511.03927*, 2015.

[BCSV06] P. Berenbrink, A. Czumaj, A. Steger, and B. Vöcking. Balanced allocations: The heavily loaded case. *SIAM Journal on Computing*, 35(6):1350–1385, 2006.

[BD13] A. Babaee and M. Draief. Distributed multivalued consensus. In *Proceedings of Computer and Information Sciences III*, pages 271–279. Springer, 2013.

[BDG+16] G. Brito, I. Dumitriu, S. Ganguly, C. Hoffman, and L. V. Tran. Recovery and rigidity in a regular stochastic block model. In *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '16)*, pages 1589–1601. SIAM, 2016.

[BDH08] M. Ben-Or, D. Dolev, and E. N. Hoch. Fast self-stabilizing byzantine tolerant digital clock synchronization. In *Proceedings of the 27th ACM Symposium on Principles of Distributed Computing*, PODC '08, pages 385–394. ACM, 2008.

[BFG03] P. Berenbrink, T. Friedetzky, and L. A. Goldberg. The natural work-stealing algorithm is stable. *SIAM Journal on Computing*, 32(5):1260–1279, 2003.

[BFGK16] P. Berenbrink, T. Friedetzky, G. Giakkoupis, and P. Kling. Efficient plurality consensus, or: the benefits of cleaning up from time to time. In I. Chatzigiannakis, M. Mitzenmacher, Y. Rabani, and D. Sangiorgi, editors, *43rd International Colloquium on Automata, Languages, and Programming (ICALP '16)*,

volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 136:1–136:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016.

[BFK+16] P. Berenbrink, T. Friedetzky, P. Kling, F. Mallmann-Trenn, L. Nagel, and C. Wastell. Self-stabilizing balls & bins in batches. In *In Proceedings of the 35th Annual ACM Symposium on Principles of Distributed Computing*, PODC '16, pages 83–92. ACM, 2016.

[BGPS06] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, pages 2508–2530, 2006.

[BKN17] L. Boczkowski, A. Korman, and E. Natale. Minimizing message size in stochastic communication patterns: Fast self-stabilizing protocols with 3 bits. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'17)*, pages 2540–2559. SIAM, 2017.

[BKR+01] A. Borodin, J. Kleinberg, P. Raghavan, M. Sudan, and D. P. Williamson. Adversarial queuing theory. *Journal of the ACM*, 48(1):13–38, 2001.

[BKSS13] P. Berenbrink, K. Khodamoradi, T. Sauerwald, and A. Stauffer. Balls-into-bins with nearly optimal load distribution. In *Proceedings of the 25th Annual ACM Symposium on Parallelism in Algorithms and Architectures (SPAA'13)*, pages 326–335. ACM, 2013.

[BLM15] C. Bordenave, M. Lelarge, and L. Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs. In *Proceedings of 56rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'15)*, pages 1347–1357. IEEE, 2015.

[Bop87] R. B Boppana. Eigenvalues and graph bisection: An average-case analysis. In *Proceedings of the 28th Annual IEEE Symposium on Foundations of Computer Science (FOCS'87)*, pages 280–285. IEEE, 1987.

[Bor15a] C. Bordenave. Personal communication, 2015.

[Bor15b] C. Bordenave. A new proof of friedman's second eigenvalue theorem and its extension to random lifts. *arXiv preprint arXiv:1502.04482*, 2015.

[BSDDS10] O. Ben-Shahar, S. Dolev, A. Dolgin, and M. Segal. Direction election in flocking swarms. In *Proceedings of the 6th International Workshop on Foundations of Mobile Computing (DIALM-POMC'10)*, pages 73–80. ACM, 2010.

[BTV09] F. Bénézit, P. Thiran, and M. Vetterli. Interval consensus: From quantized gossip to voting. In *Proceedings of the 34th*

IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '09, pages 3661–3664. IEEE Computer Society, 2009.

[Car04] M. C. Carroll. The complement system in regulation of adaptive immunity. *Nature immunology*, 5(10):981–986, 2004.

[CCD+13] A. Clementi, P. Crescenzi, C. Doerr, P. Fraigniaud, M. Isopi, A. Panconesi, F. Pasquale, and R. Silvestri. Rumor spreading in random evolving graphs. In *Proceedings of 21st European Symposium on Algorithms*, ESA '13, pages 325–336. Springer, 2013.

[CCDS14] H. Chen, R. Cummings, D. Doty, and D. Soloveichik. Speed faults in computation by chemical reaction networks. *In Distributed Computing*, page 16–30, 2014.

[CCN12] L. Cardelli and A. Csikász-Nagy. The cell cycle switch computes approximate majority. *Scientific Reports*, Vol. 2, 2012.

[CDIG+13] A. Clementi, M. Di Ianni, G. Gambosi, E. Natale, and R. Silvestri. Distributed community detection in dynamic graphs. In *Revised Selected Papers of the 20th International Colloquium on Structural Information and Communication Complexity*, volume 8179 of *SIROCCO '13*, pages 1–12. Springer-Verlag New York, Inc., 2013.

[CDIG+15] A. Clementi, M. Di Ianni, G. Gambosi, E. Natale, and R. Silvestri. Distributed community detection in dynamic graphs. *Theoretical Computer Science*, 584:19–41, 2015.

[CDRR16] S. Cannon, J. J. Daymude, D. Randall, and A. W. Richa. A markov chain algorithm for compression in self-organizing particle systems. In *Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing*, PODC '16, pages 279–288. ACM, 2016.

[CER14] C. Cooper, R. Elsasser, and T. Radzik. The power of two choices in distributed voting. In *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming (ICALP'14)*, volume 8573 of *LNCS*, pages 435–446. Springer, 2014.

[CG12] G. Cordasco and L. Gargano. Label propagation algorithm: a semi-synchronous approach. *International Journal of Social Network Mining*, 1(1):3–26, 2012.

[Cha09] B. Chazelle. Natural algorithms. In *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, (SODA '09)*, pages 422–431. SIAM, 2009.

[CHHKM12] K. Censor-Hillel, B. Haeupler, J. Kelner, and P. Maymounkov. Global computation in a poorly connected world: Fast rumor spreading with no dependence on conductance. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 961–970. ACM, 2012.

[Chu96] F. R. K. Chung. Laplacians of graphs and Cheeger's inequalities. In *Combinatorics, Paul Erdös is eighty*, volume 2, pages 157–172. János Bolyai Mathematical Society, 1996.

[CKFL05] I. D. Couzin, J. Krause, N. R. Franks, and S. A. Levin. Effective leadership and decision making in animal groups on the move. *Nature 433*, pages 513–516, 2005.

[CKW16] L. Cardelli, M. Z. Kwiatkowska, and M. Whitby. *Chemical Reaction Network Designs for Asynchronous Logic Circuits*, pages 67–81. Springer, 2016.

[CLP11] F. Chierichetti, S. Lattanzi, and A. Panconesi. Rumor spreading in social networks. *Theoretical Computer Science*, 412(24):2602–2610, 2011. (Preliminary version in ICALP'09).

[CMM+10] A. Clementi, C. Macci, A. Monti, F. Pasquale, and R. Silvestri. Flooding time of edge-markovian evolving graphs. *SIAM Journal on Discrete Mathematics*, 24(4):1694–1712, 2010.

[CO05] A. Coja-Oghlan. *Spectral techniques, semidefinite programs, and random graphs*. Habilitation thesis, Humboldt University Berlin, 2005.

[CO10] A. Coja-Oghlan. Graph partitioning via adaptive spectral techniques. *Combinatorics, Probability and Computing*, 19(02):227–284, 2010.

[Coo11] C. Cooper. Random walks, interacting particles, dynamic networks: Randomness can be helpful. In *Proceedings of the 37th International Colloquium on Structural Information and Communication Complexity (SIROCCO'11)*, pages 1–14. Springer, 2011.

[CSWB09] M. Cook, D. Soloveichik, E. Winfree, and J. Bruck. Programmability of chemical reaction networks. In *Algorithmic Bioprocesses*, pages 543–584. Springer, 2009.

[DDG+14] Z. Derakhshandeh, S. Dolev, R. Gmyr, A. W. Richa, C. Scheideler, and T. Strothmann. Brief announcement: Amoebot – a new model for programmable matter. In *Proceedings of the 26th ACM Symposium on Parallelism in Algorithms and Architectures (SPAA'14)*, pages 220–222. ACM, 2014.

[DeG74] M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[DF89] M. E. Dyer and A. M. Frieze. The solution of some random NP-hard problems in polynomial expected time. *Journal of Algorithms*, 10(4):451–489, 1989.

[DF11] B. Doerr and M. Fouz. Asymptotically optimal randomized rumor spreading. *Electronic Notes in Discrete Mathematics*, 38:297–302, 2011.

[DGH+87] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry. Epidemic

algorithms for replicated database maintenance. In *Proceedings of the 6th Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC'12)*, pages 1–12. ACM, 1987.

[DGH+88] A. J. Demers, D. H. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. E. Sturgis, D. C. Swinehart, and D. B. Terry. Epidemic algorithms for replicated database maintenance. *Operating Systems Review*, 22(1):8–32, 1988.

[DGM+10] M. Dietzfelbinger, A. Goerdt, M. Mitzenmacher, A. Montanari, R. Pagh, and M. Rink. Tight thresholds for cuckoo hashing via xorsat. In *Proceedings of the 37th International Colloquium on Automata, Languages, and Programming (ICALP'10)*, volume 6198 of *LNCS*, pages 213–225. Springer, 2010.

[DGM+11] B. Doerr, L. A. Goldberg, L. Minder, T. Sauerwald, and C. Scheideler. Stabilizing consensus with the power of two choices. In *Proceedings of the 23rd Annual ACM Symposium on Parallelism in Algorithms and Architectures (SPAA'11)*, pages 149–158. ACM, 2011.

[DH07] D. Dolev and E. N. Hoch. On self-stabilizing synchronous actions despite byzantine attacks. In *Proceedings of the International Symposium on Distributed Computing (DISC'07)*, pages 193–207, 2007.

[Dij74] E. W. Dijkstra. Self-stabilizing systems in spite of distributed control. *Communications of the ACM*, 17(11):643–644, 1974.

[DK70] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

[DKMZ11] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.

[Dol97] S. Dolev. Possible and impossible self-stabilizing digital clock synchronizationin general graphs. *Real-Time Systems*, 12(1):95–107, 1997.

[Dol00] S. Dolev. *Self-stabilization*. MIT press, 2000.

[Doo53] J. L. Doob. *Stochastic Processes*. John Wiley & Sons Inc., 1953.

[Dot14] D. Doty. Timing in chemical reaction networks. In *Proceedings of 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'14)*, pages 772–784. SIAM, 2014.

[DP09] D. P. Dubhashi and A. Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[DR98] D. Dubhashi and D. Ranjan. Balls and bins: A study in negative dependence. *Random Structures and Algorithms*, 13(2):99–124, 1998.

[DS15] D. Doty and D. Soloveichik. Stable leader election in population protocols requires linear time. *arXiv preprint arXiv:1502.04246*, 2015.

[DSMP12] A. Das Sarma, A. R. Molla, and G. Pandurangan. Near-optimal random walk sampling in distributed networks. In *Proceedings of the 31th IEEE Conference on Computer Communications (INFOCOM'12)*, 2012.

[DV12] M. Draief and M. Vojnovic. Convergence speed of binary interval consensus. *SIAM Journal on Control and Optimization*, 50(3):1087–1109, 2012.

[DW04] S. Dolev and J. L. Welch. Self-stabilizing clock synchronization in the presence of byzantine faults. *Journal of the ACM*, 51(5):780–799, 2004.

[EGK11] A. El Gamal and Y. Kim. *Network information theory*. Cambridge university press, 2011.

[EK10] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets*. Cambridge University Press, 2010.

[EK15] R. Elsässer and D. Kaaser. On the influence of graph density on randomized gossiping. *Proceedings of the 29th IEEE International Parallel & Distributed Processing Symposium (IPDPS'15)*, pages 521–531, 2015.

[ES09] R. Elsässer and T. Sauerwald. On the runtime and robustness of randomized broadcasting. *Theoretical Computer Science*, 410(36):3414–3427, 2009.

[EW13] Y. Emek and R. Wattenhofer. Stone age distributed computing. In *Proceedings of the 2013 ACM Symposium on Principles of Distributed Computing (PODC '13)*, pages 137–146. Springer, 2013.

[FHK14] O. Feinerman, B. Haeupler, and A. Korman. Breathe before speaking: Efficient information dissemination despite noisy, limited and anonymous communication. In *Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC '14)*, pages 114–123. ACM, 2014.

[FHK15] O. Feinerman, B. Haeupler, and A. Korman. Breathe before speaking: efficient information dissemination despite noisy, limited and anonymous communication. *Distributed Computing*, pages 1–17, 2015.

[FJ90] N. E. Friedkin and E. C. Johnsen. Social influence and opinions. *The Journal of Mathematical Sociology*, 15(3-4):193–206, 1990.

[FJT+10] O. Feinerman, G. Jentsch, K. E. Tkach, J. W. Coward, M. M. Hathorn, M. W. Sneddon, T. Emonet, K. A. Smith, and

G. Altan-Bonnet. Single-cell quantification of il-2 response by effector and regulatory t cells reveals critical plasticity in immune response. *Molecular systems biology*, 6(1):437, 2010.

[FK14] J. Friedman and D. E. Kohler. The relativized second eigenvalue conjecture of alon. *arXiv preprint arXiv:1403.3462*, 2014.

[FK15] O. Feinerman and A. Korman. Clock synchronization and estimation in highly dynamic networks: An information theoretic approach. In *Proceedings of the International Colloquium on Structural Information and Communication Complexity (SIROCCO'15)*, pages 16–30, 2015.

[FKP11] P. Fraigniaud, A. Korman, and D. Peleg. Local distributed decision. In *Proceedings of the 52th Annual IEEE Symposium on Foundations of Computer Science (FOCS'11)*, 2011.

[FN16] P. Fraigniaud and E. Natale. Noisy rumor spreading and plurality consensus. In *Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing (PODC'16)*, pages 127–136. ACM, 2016.

[FPM+02] N. R. Franks, S. C. Pratt, E. B. Mallon, N. F. Britton, and D. J. T. Sumpter. Information flow, opinion polling and collective intelligence in house–hunting social insects. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 357(1427):1567–1583, 2002.

[GG] O. Gurel-Gurevich. Expectation of square root of binomial r.v. MathOverflow. URL:http://mathoverflow.net/q/121424 (version: 2013-02-10).

[GJS76] M. R. Garey, D. S. Johnson, and L. Stockmeyer. Some simplified np-complete graph problems. *Theoretical computer science*, 1(3):237–267, 1976.

[GK10] S. Gilbert and D. Kowalski. Distributed agreement with optimal communication complexity. In *Proceedings of 21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'10)*, pages 965–977. SIAM, 2010.

[GS14] David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science (ITCS'14)*, pages 369–376. ACM, 2014.

[Haj82] B. Hajek. Hitting-time and occupation-time bounds implied by drift analysis with applications. *Advances in Applied probability*, 14(3):502–525, 1982.

[Her00] T. Herman. Phase clocks for transient fault repair. *IEEE Transactions on Parallel and Distributed Systems*, 11(10):1048–1057, 2000.

[HLL83] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[HLW06] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[HM85] R. D. Harkness and N. G. Maroudas. Central place foraging by an ant (*cataglyphis bicolor* fab.): a model of searching. *Animal Behavior 33(3)*, pages 916–928, 1985.

[HP01] Y. Hassin and D. Peleg. Distributed probabilistic polling and applications to proportionate agreement. *Information and Computation*, 171(2):248–268, 2001.

[HPP+12] B. Haeupler, G. Pandurangan, D. Peleg, R. Rajaraman, and Z. Sun. Discovery through gossip. In *Proceedings of the 24th ACM Symposium on Parallelism in Algorithms and Architectures (SPAA'12)*. ACM, 2012.

[HPP+16] B. Haeupler, G. Pandurangan, D. Peleg, R. Rajaraman, and Z. Sun. Discovery through gossip. *Random Structures & Algorithms*, 48(3):565–587, 2016.

[HW90] B. Hölldobler and E. O. Wilson. *The ants*. Harvard University Press, 1990.

[HW92] J. M. Harrison and R. J. Williams. Brownian models of feedforward queueing networks: Quasireversibility and product form solutions. *The Annals of Applied Probability*, 2(2):263–293, 1992.

[HW09] B. Hölldobler and E. O. Wilson. *The superorganism: the beauty, elegance, and strangeness of insect societies*. WW Norton & Company, 2009.

[IJ90] A. Israeli and M. Jalfon. Token management schemes and random walks yield self-stabilizing mutual exclusion. In *Proceedings of the 9th annual ACM Symposium on principles of Distributed Computing (PODC'90)*, pages 119–131. ACM, 1990.

[IKOY02] S. Ikeda, I. Kubo, N. Okumoto, and M. Yamashita. Fair circulation of a token. *IEEE Transactions on Parallel and Distributed Systems*, 13(4):367–372, 2002.

[JKV12] K. Jung, B. Y. Kim, and M. Vojnovic. Distributed ranking in networks with limited memory and communication. In *Proceedings of the 2012 IEEE International Symposium on Information Theory (ISIT'12)*. IEEE, 2012.

[JS98] M. Jerrum and G. B. Sorkin. The metropolis algorithm for graph bisection. *Discrete Applied Mathematics*, 82(1):155–175, 1998.

[KDG03] D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *Proceedings of 43rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'03)*, pages 482–491. IEEE, 2003.

[KK13] A. Kravchik and S. Kutten. Time optimal synchronous self stabilizing spanning tree. In *Proceedings of the International*

Symposium on Distributed Computing (DISC'13), pages 91–105, 2013.

[Kle99] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[KLMadH96] R. M. Karp, M. Luby, and F. Meyer auf der Heide. Efficient pram simulation on a distributed memory machine. *Algorithmica*, 16(4-5):517–542, 1996.

[KM04] D. Kempe and F. McSherry. A decentralized algorithm for spectral analysis. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC'04)*, pages 561–568. ACM, 2004.

[KMM+13] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013.

[KMTN15] D. Kaaser, F. Mallmann-Trenn, and E. Natale. On the voting time of the deterministic majority process. In *41st International Symposium on Mathematical Foundations of Computer Science (MFCS 2016)*. Springer, 2015.

[KPS13] K. Kothapalli, S. V. Pemmaraju, and V. Sardeshmukh. On the analysis of a label propagation algorithm for community detection. In *International Conference on Distributed Computing and Networking (ICDCN'13)*, pages 255–269. Springer, 2013.

[KSSV00] R. Karp, C. Schindelhauer, S. Shenker, and B. Vocking. Randomized rumor spreading. In *Proceedings of the 41th Annual IEEE Symposium on Foundations of Computer Science (FOCS'00)*, pages 565–574. IEEE, 2000.

[KT08] M. Kearns and J. Tan. Biased voting and the democratic primary problem. In *Proceedings of the 4th Workshop on Internet and Network Economics (WINE'08)*, volume 5385 of *Lectures Notes in Computer Science*, pages 639–652. Springer, 2008.

[Lam78] L. Lamport. Time, clocks, and the ordering of events in a distributed system. *Communications of the ACM*, 21(7):558–565, 1978.

[Lam85] L. Lamport. Solved problems, unsolved problems and nonproblems in concurrency. *ACM SIGOPS Operating Systems Review*, 19(4):34–44, 1985.

[Lan50] C. Lanczos. *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*. United States Government Press Office Los Angeles, CA, 1950.

[LB95] M. W. S. Land and R. K. Belew. No two-state ca for density classification exists. *Physical Review Letters*, 74(25):5148–5150, 1995.

[LHLC09] I. X. Y. Leung, P. Hui, P. Lio, and J. Crowcroft. Towards real-time community detection in large networks. *Physical Review E*, 79(6):066107, 2009.

[Lig12] T. Liggett. *Interacting particle systems*, volume 276. Springer, 2012.

[LLSW10] C. Lenzen, T. Locher, P. Sommer, and R. Wattenhofer. Clock synchronization: Open problems in theory and practice. In *International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM'10)*, pages 61–70. Springer, 2010.

[LLW10] C. Lenzen, T. Locher, and R. Wattenhofer. Tight bounds for clock synchronization. *Journal of the ACM*, 57(2), 2010.

[LM10] X. Liu and T. Murata. Advanced modularity-specialized label propagation algorithm for detecting communities in networks. *Physica A: Statistical Mechanics and its Applications*, 389(7):1493–1500, 2010.

[LPW09] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2009.

[LR15] C. Lenzen and J. Rybicki. Efficient counting with optimal resilience. In *Proceedings of the International Symposium on Distributed Computing (DISC'15)*, pages 16–30. Springer, 2015.

[LRS15] C. Lenzen, J. Rybicki, and J. Suomela. Towards optimal synchronous counting. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing (PODC'15)*, pages 441–450. ACM, 2015.

[LV15] C. M. Le and R. Vershynin. Concentration and regularization of random graphs. *arXiv preprint arXiv:1506.00669*, 2015.

[Lyn96] N. A. Lynch. *Distributed algorithms*. Morgan Kaufmann, 1996.

[Mac03] D. J. C. MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[Mas14] L. Massoulie. Community detection thresholds and the weak ramanujan property. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC'14)*, pages 694–703. ACM, 2014.

[McD98] C. McDiarmid. *Concentration*, pages 195–248. Springer, 1998.

[McS01] F. McSherry. Spectral partitioning of random graphs. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS'01)*, pages 529–537. IEEE, 2001.

[Mit01] M. Mitzenmacher. The power of two choices in randomized load balancing. *IEEE Transactions on Parallel and Distributed Systems*, 12(10):1094–1104, 2001.

[MK07] J. M. Mooij and H. J. Kappen. Sufficient conditions for convergence of the sum–product algorithm. *IEEE Transactions on Information Theory*, 53(12):4422–4437, 2007.

[MNRS14] G. B. Mertzios, S. E. Nikoletseas, C. Raptopoulos, and P. G. Spirakis. Determining majority in networks with local interactions and very small local memory. In *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming (ICALP'14)*, 2014.

[MNS13] E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. *arXiv preprint arXiv:1311.4115*, 2013.

[MNS14] E. Mossel, J. Neeman, and A. Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3-4):431–461, 2014.

[MNT14] E. Mossel, J. Neeman, and O. Tamuz. Majority dynamics and aggregation of information in social networks. *Autonomous Agents and Multi-Agent Systems*, 28(3):408–429, 2014.

[Mou14] N. Mousavi. How tight is chernoff bound?, 2014. url: https://ece.uwaterloo.ca/ nmousavi/Papers/Chernoff-Tightness.pdf.

[MPS02] M. Mitzenmacher, B. Prabhakar, and D. Shah. Load balancing with memory. In *Proceedings of the 43th Annual IEEE Symposium on Foundations of Computer Science (FOCS'02)*, pages 799–808. IEEE, 2002.

[MRSDZ11] Y. Métivier, J. M. Robson, N. Saheb-Djahromi, and A. Zemmari. An optimal bit complexity randomized distributed mis algorithm. *Distributed Computing*, 23(5-6):331–340, 2011.

[MS10] E. Mossel and G. Schoenebeck. Reaching consensus on social networks. In *Proceedings of the 2nd Innovations in Computer Science (ITCS'10)*, pages 214–229, 2010.

[MU05] M. Mitzenmacher and E. Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge University Press, 2005.

[New02] M. E. J. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(1), 2002.

[OT09] A. Olshevsky and J. N. Tsitsiklis. Convergence speed in distributed consensus and averaging. *SIAM Journal on Control and Optimization*, 48(1):33–55, 2009.

[Pel00] D. Peleg. Distributed computing. *SIAM Monographs on discrete mathematics and applications*, 5, 2000.

[Pel02] D. Peleg. Local majorities, coalitions and monopolies in graphs: a review. *Theoretical Computer Science*, 282(2):231–257, 2002.

[Pit87] B. Pittel. On spreading a rumor. *SIAM Journal on Applied Mathematics*, 47(1):213–223, 1987.

[PSL80] M. Pease, R. Shostak, and L. Lamport. Reaching agreement in the presence of faults. *Journal of the ACM*, 27(2):228–234, 1980.

[PVV09] E. Perron, D. Vasudevan, and M. Vojnovic. Using three states for binary consensus on complete graphs. In *Proceedings of the 28th IEEE Conference on Computer Communications (INFOCOM'09)*, pages 2527–1535. IEEE, 2009.

[Rab83] M. O. Rabin. Randomized byzantine generals. In *Proceedings of the 24th Annual Symposium on Foundations of Computer Science (SFCS'83)*, pages 403–409. IEEE, 1983.

[RAK07] U. N. Raghavan, R. Albert, and S. Kumara. Near linear time algorithm to detect community structures in large—scale networks. *Physical Review E*, 76(3):036106, 2007.

[REF13] N. Razin, J. P. Eckmann, and O. Feinerman. Desert ants achieve reliable recruitment across noisy interactions. *Journal of the Royal Society Interface; 10(20170079)*, 2013.

[Reu16] B. Reus. *Molecular Computing*, pages 299–316. Springer, 2016.

[RM08] Y. Ruan and Y. Mostofi. Binary consensus with soft information processing in cooperative networks. In *Proceedings of the 47th IEEE Conference on Decision and Control (CDC'08)*, pages 3613–3619. IEEE, 2008.

[Rob55] H. Robbins. A remark on stirling's formula. *The American Mathematical Monthly*, 62(1):26–29, 1955.

[Rob96] G. Roberts. Why individual vigilance increases as group size increases. *Animal Behaviour 51*, pages 1077–1086, 1996.

[RS98] M. Raab and A. Steger. "balls into bins"—a simple and tight analysis. In *Proceedings of the 2nd International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM'98)*, pages 159–170. Springer, 1998.

[San06] N. Santoro. *Design and analysis of distributed algorithms*. John Wiley & Sons, 2006.

[Sha09] D. Shah. *Gossip algorithms*. Now Publishers Inc, 2009.

[SKJ+08] D. J. T. Sumpter, J. Krause, R. James, I. D. Couzin, and A. J. W. Ward. Consensus decision making by fish. *Current biology : CB*, 18 22:1773–7, 2008.

[Spr13] Springer. *Synchronous counting and computational algorithm design*, 2013.

[SS90] P. Stewart and J. Sun. *Matrix Perturbation Theory*. Academic Press, 1990.

[Suo13] J. Suomela. Survey of local algorithms. *ACM Computing Surveys*, 45(2):24:1–24:40, 2013.

[Tsi84] J. N. Tsitsiklis. Problems in decentralized decision making and computation. Technical report, DTIC Document, 1984.

[Vöc03] B. Vöcking. How asymmetry helps load balancing. *Journal of the ACM*, 50(4):568–589, 2003.

[WB05] C. M. Waters and B. L. Bassler. Quorum sensing: cell-to-cell communication in bacteria. *Annual review of cell and developmental biology*, 21:319–346, 2005.

[Wei00] Y. Weiss. Correctness of local probability propagation in graphical models with loops. *Neural computation*, 12(1):1–41, 2000.

[Wil91] D. Williams. *Probability with Martingales*. Cambridge University Press, 1991.

[Wil92] G. S. Wilkinson. Information transfer at evening bat colonies. *Animal Behaviour 44*, pages 501–518, 1992.

[Wol02] S. Wolfram. *A New Kind of Science*, volume 5. Wolfram media Champaign, 2002.

[XBK07] L. Xiao, S. Boyd, and S. Kim. Distributed average consensus with least-mean-square deviation. *Journal of Parallel and Distributed Computing*, 67(1):33–46, 2007.

[YOA+13] E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, and A. Scaglione. Binary opinion dynamics with stubborn agents. *ACM Transactions on Economics and Computation*, 4(1):Article n. 19, 2013.