HAL
open science

Social network dynamics of face-to-face interactions

Kun Zhao, Juliette Stehle, Ginestra Bianconi, Alain Barrat

To cite this version:

Kun Zhao, Juliette Stehle, Ginestra Bianconi, Alain Barrat. Social network dynamics of face-to-face interactions. Physical Review E : Statistical, Nonlinear, and Soft Matter Physics, 2011, 83, pp.056109. 10.1103/PhysRevE.83.056109 . hal-00565802

HAL Id: hal-00565802

https://hal.science/hal-00565802v1

Submitted on 7 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

arXiv:1102.2423v2 [physics.soc-ph] 17 May 2011

Social network dynamics of face-to-face interactions

Kun Zhao,[1] Juliette Stehlé,[2] Ginestra Bianconi,[1] and Alain Barrat[2,3]

[1] *Department of Physics, Northeastern University, Boston 02115 MA, USA*
[2] *Centre de Physique Théorique, CNRS (UMR 6207) et Université d'Aix-Marseille, Campus de Luminy, Case 907, F-13288 Marseille cedex 9, France*
[3] *Complex Networks and Systems Group, Institute for Scientific Interchange (ISI), Torino 10133, Italy*

The recent availability of data describing social networks is changing our understanding of the "microscopic structure" of a social tie. A social tie indeed is an aggregated outcome of many social interactions such as face-to-face conversations or phone calls. Analysis of data on face-to-face interactions shows that such events, as many other human activities, are bursty, with very heterogeneous durations. In this paper we present a model for social interactions at short time scales, aimed at describing contexts such as conference venues in which individuals interact in small groups. We present a detailed analytical and numerical study of the model's dynamical properties, and show that it reproduces important features of empirical data. The model allows for many generalizations toward an increasingly realistic description of social interactions. In particular in this paper we investigate the case where the agents have intrinsic heterogeneities in their social behavior, or where dynamic variations of the local number of individuals are included. Finally we propose this model as a very flexible framework to investigate how dynamical processes unfold in social networks.

PACS numbers: 89.75.-k,64.60.aq,89.65.-s,89.20.-a

I. INTRODUCTION

In the last decade complexity theory has greatly advanced thanks to the availability of extensive data on a wide variety of networked systems. Empirical studies have uncovered the presence of ubiquitous features in complex networks, such as the small-world property, or strong heterogeneities in the topological structure, revealed for instance by broad degree distributions [1–5]. These findings have deeply affected our understanding of self-organized networks and have been used for the investigation, characterization and modeling of many different systems such as infrastructure or biological and social networks.

Many works have studied the influence of complex network topologies observed in real networks on the dynamical phenomena that unfold on them [6, 7]. While these studies have mostly focused on networks considered as static objects with a fixed topology, networks' structures may in principle evolve, links may appear and disappear. A first approach consists of assuming that links are created and annihilated at a constant rate, independently of the dynamical process that takes place on them [8–10]. Networks can however display more interesting properties such as an adaptative behavior, in which the dynamics on the network and of the network are related by feedback effects [11–17]. Moreover, empirical investigations have shown that link durations can significantly deviate from a Poisson process [18–23].

Social networks [24, 25] are prominent examples of evolving networks. Social relationships are indeed continuously changing, possibly in a way correlated with the dynamical processes taking place during social interactions (such as epidemic spreading or opinion dynamics). Consequently, a number of works have been devoted to modeling the dynamics of social interactions. Issues investigated in this context are in particular community formation [26–28] and the evolution of adaptive dynamics of opinions and social ties through schematic models in which links can disappear or be rewired at random [11–17]. Moreover, social networks evolve on many different timescales. The static representation of social ties hide indeed dynamical sequences of events such as face-to-face interactions, phone calls or email exchanges, and can be measured by aggregating fast social interactions over a certain period of time.

Recently, technological advances have made possible the access to data sets that give new insights into such link internal dynamics, characterized by sequences of events of different durations. Traces of human behavior are often unwittingly recorded in a variety of contexts (financial transactions, phone calls, mobility patterns, purchases using credit cards, etc.). Data have been gathered and analyzed about the mobility patterns inside a city [29], between cities [30], as well as at the country and at worldwide levels [20, 31–33]. At a more detailed level, mobile devices such as cell phones make it possible to investigate individual mobility patterns and their predictability [34, 35]. Mobile devices and wearable sensors using Bluetooth and Wifi technologies give access to proximity patterns of pairs of individuals [18, 36–39], and even face-to-face presence can be resolved with high spatial and temporal resolution [21–23, 40]. Finally, on-line interactions occurring between individuals can be monitored by logging instant messaging or email exchange [41–47].

The combination of these technological advances and of heterogeneous data sources allows researchers to gather longitudinal data that have been traditionally scarce in social network analysis [48, 49]. Analysis of such data sets

has clearly shown the bursty nature of many human and social activities, revealing the inadequacy of many traditional frameworks that posit Poisson distributed processes. In particular, the durations of "contacts" between individuals, as defined by the proximity of these individuals, display broad distributions, as well as the time intervals between successive contacts [18, 19, 21–23, 39]. Burstiness of interactions has strong consequences on dynamical processes [23, 50–54], and should therefore be correctly taken into account when modeling the interaction networks. New frameworks are therefore needed, which integrate the bursty character of human interactions and behaviors into dynamic network models.

While a lot of modeling efforts have been devoted to static networks, the development of models of dynamic networks has indeed until recently attracted less attention [17, 19, 20, 55, 56]. In a recent paper [56], we have presented an agent-based modeling framework to describe how individuals interact at short times scales in venues such as social gatherings (e.g., scientific conferences). The model is based on a reinforcement dynamics (in the spirit of the preferential attachment in complex networks [57] and of Hebbian learning) which might be responsible for the bursty dynamics of social face-to-face interactions. The proposed mechanism implies that *the longer an agent is interacting in a group, the smaller is the probability that he/she will leave the group; the longer an agent is isolated the smaller is the probability that he/she will form a new group.*

In the present paper, we present an extensive characterization of the model's properties, and show that it reproduces important features of empirical data on social interactions at short timescales. We characterize the rich phase diagram of the model, which includes stationary and non-stationary regions. The analysis of the dynamical properties of the model shows that it yields stationary broad distribution of group lifetimes even if the underlying dynamics is non-stationary, as also found in empirical data. In order to illustrate the model's versatility, we give two examples of how it can be extended to more realistic cases: in the first example, agents can have heterogeneous propensities to form group with others; in the second example, we introduce the possibility of a varying population, where the number of individuals can be an arbitrary function, or extracted from empirical data. In the two cases we show how properties very close to the ones of real-world data sets can be obtained. The proposed model is easily implementable, uses simple but realistic mechanisms, reproduces a certain number of empirical facts, and is amenable to further refinements. It can therefore be used to produce artificial data sets of bursty interaction networks on which dynamical phenomena can be simulated and studied.

The paper is organized as follows. In Sec. II, we review the main properties of a representative empirical data set describing the face-to-face proximity of individuals in social gatherings. Section. III is devoted to the definition of the modeling framework and to the analytical and numerical study of the simplest versions of the model. Section. IV is devoted to two extensions of the model. We outline some conclusions and perspectives in Sec. V.

II. EMPIRICAL DATA

The infrastructure developed by the SocioPatterns project, described in [21–23], has yielded measurements about the face-to-face proximity of individuals in different types of social contexts (hospital, primary school, scientific conferences, museum), with a fine grained time resolution. The infrastructure is currently based on radio frequency identification devices (RFID). Individuals participating to the data collection are asked to wear small RFID tags on their chests (as a conference badge), that emit radio packets at very low power. The parameters of the infrastructure are tuned so that the tags can exchange radio packets only when the individuals wearing them face each other at close range (about 1 to 1.5 m), and so that face-to-face proximity events ("contacts") can be assessed with very high accuracy with a time resolution of 20s. We present here for the sake of completeness some characteristics of the data collected during the 6th European Semantic Web Conference (ESWC, Heraklion, Greece) in 2009. Analysis of other data sets can be found in Refs. [21–23], with very similar features.

Face-to-face proximity patterns have been collected for 175 voluntary participants (among the 305 conference attendants) over 3 days. 14520 contact events have been registered, corresponding to 27.3 contacts per individual per day. The contact durations display a very broad distribution, close to a power law, (shown in Fig. 1): the average duration is of 46 s, but long-lasting contacts are as well observed, and no characteristic contact timescale can be extracted from the data. Figure 1 also displays a quantity of interest, namely the distribution of time intervals between the start of two consecutive contacts of a given individual A with two distinct persons B and C. In other words, if A starts a contact with B at time t_{AB} , and then starts a different contact with C at t_{AC}, the inter-contact interval is defined as $\tau = t_{AC} - t_{AB}$. These time intervals constrain causal processes such as information diffusion or epidemic spreading, as they determine the timescale after which an individual receiving some information or disease is able to propagate it to another individual. As shown in Fig. 1, broad distributions are also found in this case. We also show in Fig. 2 the distributions of lifetimes of groups of size $n + 1$ ($n = 0$ corresponds to an isolated person, $n = 1$ to a pair, etc). All these distributions are broad, compatible with power-law shapes, and become narrower for increasing n (larger groups are less stable than smaller ones).

The burstiness of the contact pattern revealed by the broad distribution of contact durations has consequences also on aggregated views of the dynamical contact network. Aggregated contact networks over a given time

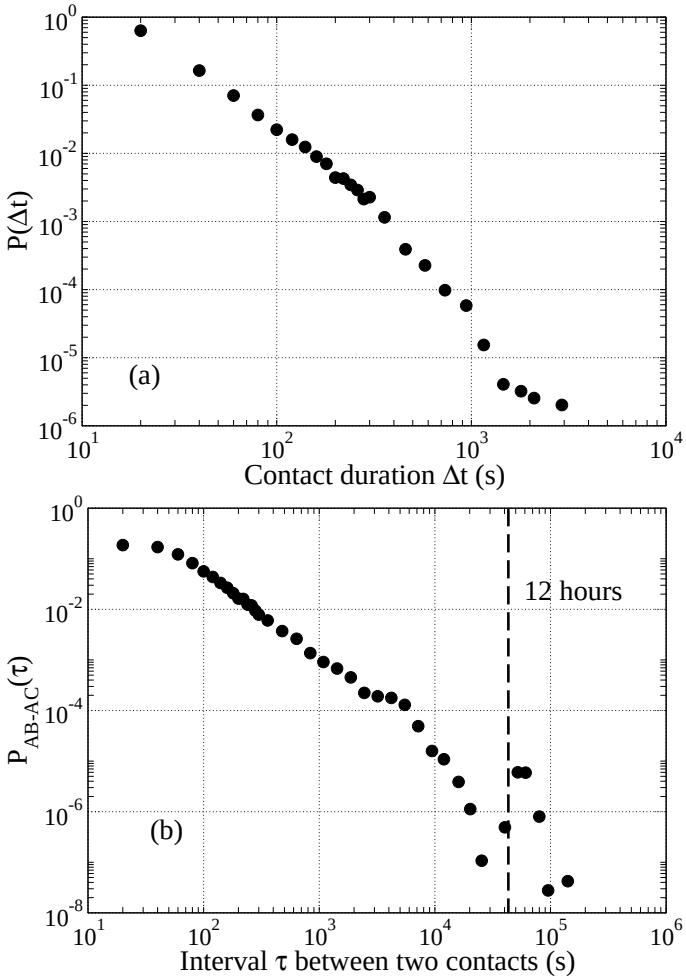

FIG. 1: Distribution of contact durations (a) and of time intervals between two contacts involving a common individual (b).

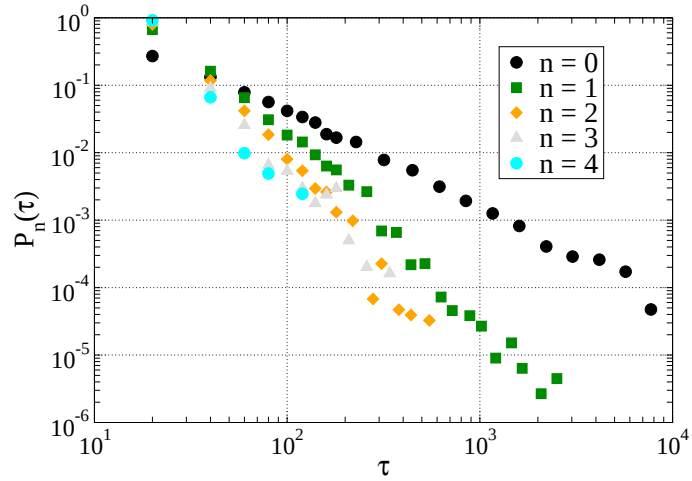

FIG. 2: (Color online) Distributions $P_n(\tau)$ of the durations

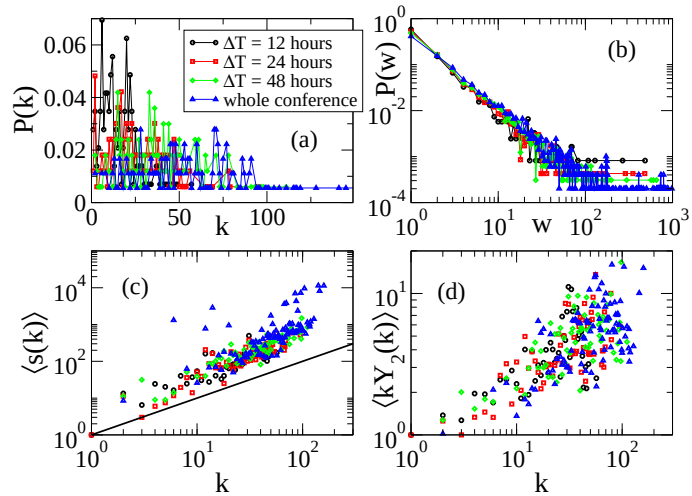

FIG. 3: (Color online) Aggregated networks' characteristics for the empirical ESWC data: (a) degree distribution, (b) weight distribution, (c) average strength of nodes of degree k, vs k, and (d) average Herfindahl-Hirschman index of nodes of degree k, vs k.

window are defined as follows: each node corresponds to an individual, and an edge is drawn between two nodes if at least one contact event has been registered during the time window between the two corresponding individuals. Each edge is weighted by the sum of the contact durations between these individuals during this time window. As shown in Fig. 3 (top right), the distributions of such weights are broad, independent of the time window considered (see also [22, 23]).

Figure 3 also displays other characteristics of aggregated networks constructed with time windows of different lengths. As also found in [22, 23] for other deployments of the same infrastructure, the distribution of degrees (the degree gives the number of distinct individuals with whom a given individual has been in contact) are not broadly distributed. This behavior is in contrast with the degree distribution of many empirically studied social networks [6, 58, 59]. It should however be emphasized that we are here dealing with face-to-face interactions occurring in a restricted environment among a relatively small population and on short timescales, while studies such as [58, 59] are concerned with social ties created and defined over much longer times cales. Moreover, we note that a narrow degree distribution (a power law with a large exponent) has also been found in the degree distribution of networks defined by phone-calls data sets [52].

The strength s_i of a node i is defined as the sum of the weights w_{ij} between i and its neighbors j, that is,

$$s_i = \sum_j w_{ij}. \tag{1}$$

It gives the cumulated durations of the interactions of the corresponding individual. The average strength of nodes of degree k $[\langle s(k)\rangle]$ indicates how the weights are dis-

tributed. If the weights are uniformly distributed among the links of the networks, the average strength $\langle s(k)\rangle$ grows linearly with k, that is, $\langle s(k)\rangle \propto k\langle w\rangle$. On the contrary, if stronger ties are more frequently linked to highly connected agents, a superlinear behavior of $\langle s(k)\rangle$ versus k is observed [31]. The RFID data on face-to face interaction are consistent with a linear or slightly faster behavior of $\langle s(k)\rangle$ versus k [22, 23], hinting at a weak correlation between weights and degrees. Correlations between network topology and distribution of the weights have been found in various complex networks with broad degree distribution [31], and several models based on reinforcement dynamics have been proposed in this context [60–62].

Another measure of the weights' distribution is given by the Herfindahl-Hirschman index Y_2 [63, 64], also known in the physics literature under the name of "participation ratio". This index, defined as

$$Y_2(i) = \sum_{j\in\mathcal{N}(i)} \left(\frac{w_{ij}}{s_i}\right)^2, \tag{2}$$

where $\mathcal{N}(i)$ refers to the set of neighbors of i, gives a measure of the heterogeneity of the weights among the neighbors of a node. When all weights w_{ij} of the links connected to node i are equal, that is, $w_{ij} = s_i/k_i$, this index is inversely proportional to the degree k_i of node i, that is,

$$Y_2(i) = \frac{1}{k_i}. \tag{3}$$

On the contrary, when one of the links has a much larger weight than the others, $Y_2(i)$ is close to 1. The departure of $k_iY_2(i)$ from 1 thus indicates the local heterogeneity of weights around each node. Figure 3 shows that the average of this quantity over nodes of degree k [$kY_2(k)$], is larger than 1 and increases with k: each individual divides his/her time unevenly among her/his contacts, and this behavior becomes more pronounced as the number of distinct contacted persons increases.

III. A MODEL FOR SOCIAL DYNAMICS AT SHORT TIMESCALES

In this section, we present a model describing the social dynamics of N agents forming small groups of different size. In this model, each agent can interact with any other agent. This model describes therefore social interactions within a closed environment of relatively small size where agents are free to meet, such as a conference venue. We assign to each agent $i = 1, 2, \ldots, N$ a coordination number $n_i = 0, 1, 2 \ldots, N$ indicating the number of agents interacting with him/her. If an agent i has coordination number $n_i = 0$ he/she is isolated, and if $n_i = n > 0$ he/she is part of a group of $n + 1$ agents, who all interact with each other (thus forming a clique). We also assign to each agent i the temporal variable t_i indicating the last time at which his/her coordination number n_i has changed.

The dynamics of the model is as follows. Starting from random initial conditions, at each time step t the following steps are performed:

(1) An agent i is chosen randomly.

(2) The agent i updates his/her coordination number $n_i = n$ with a certain probability $p_n(t, t_i)$ that may depend on the agent's state, on the present time t, and on the last time t_i at which i's state evolved. With probability $1 - p_n(t, t_i)$, the agent does not change state.

If the coordination number n_i is updated, the action of the agent is chosen with the following rules.

(i) If the agent i is isolated, that is, $n_i = 0$, he/she starts an interaction with another isolated agent j chosen with probability proportional to $p_0(t, t_j)$. The coordination number of the agent i and of the agent j are then updated according to the rule $n_i \to 1$ and $n_j \to 1$.

(ii) If the agent i is interacting in a group, that is, $n_i = n > 0$, with probability λ the agent leaves the group and with probability $(1 - \lambda)$ he/she introduces an isolated agent in the group. If the agent i leaves the group, his/her coordination number is updated ($n_i \to 0$), as well as the coordination numbers of all the agents in the original group, that is, $n_k \to n - 1$ (for all agents k in the original group). On the contrary, if the agent i introduces another isolated agent j to the group, j is chosen with probability proportional to $p_0(t, t_j)$ and the coordination numbers of all the interacting agents are changed according to the rules $n_j \to n + 1$ and $n_k \to n + 1$ (for all k in the group).

The structure and properties of the interactions between agents depend on the choice of the probabilities p_n, which control the tendency of the agents to change their state, and on the parameter λ, which determines the tendency either to leave groups or on the contrary to make them grow. The simplest choice consists of considering constant probabilities $p_n(t, t') = p_n$: at each time, every agent has a fixed probability to form a group or split from a group. In this case the formation of the groups is a Poisson process, and the distributions of the durations of contacts between agents, or the lifetime of a group, are exponentially distributed.

As recalled in the introduction however, the distributions of the times describing human activities are typically broad [18, 22, 23, 37, 42, 46, 51], and are clearly closer to power-laws that lack a characteristic time scale than to exponentials. A possible explanation of such results is given by mechanisms in which the decisions of

the agents to form or leave a group are driven by memory effects dictated by reinforcement dynamics, that can be summarized in the following statement: *the longer an agent is interacting in a group, the smaller is the probability that he/she will leave the group; the longer an agent is isolated, the smaller is the probability that he/she will form a new group.* In particular, such reinforcement principle implies that the probabilities $p_n(t,t')$ that an agent with coordination number n changes his/her state depend on the time elapsed since his/her last change of state. A simple way to introduce this hypothesis is to consider functions $p_n(t,t') = p_n(t-t')$. Reinforcement mechanisms are then described by decreasing functions p_n. We will see in the next subsections how the evolution equations of the number of agents in each state can be written at the mean-field level for arbitrary functions p_n. Finding solutions of this set of equations is however not always possible, and we will focus on functions p_n scaling as $1/(t-t')$, for two reasons: on the one hand, it represents one of the cases that is fully amenable to analytical computations and, on the other hand, such a scaling behavior is needed to obtain power-law distributions for the contact durations and thus dynamical characteristics compatible with empirical data. We consider in particular functions given by

$$p_n(t,t') = \frac{b_n}{1+(t-t')/N}, \tag{4}$$

and, in order to reduce the number of parameters, we moreover take $b_n = b_1$ for every $n \geq 1$, indicating the fact that interacting agents change their state independently on the number of other agents n with whom they are interacting, provided that $n \geq 1$. The model's parameter are thus b_0, b_1, and λ.

A. Pairwise interactions

Let us first consider a restricted version of the model, in which the agents can only interact in pairs. This setup is obtained by setting $\lambda = 1$ and by considering initial conditions in which the agents interact at most in groups of size 2. In this case, each agent is thus assigned a variable $n_i = 0, 1$ indicating if the agent i is isolated ($n_i = 0$) or interacting with another agent ($n_i = 1$).

As in the analysis of empirical data, the most immediate quantities of interest concern the time spent by agents in each state, the duration of contacts between two agents, and the time intervals between successive contacts of an agent. To gain insight into these temporal properties of the system, we can write rate equations for the evolution of the numbers $N_n(t,t')$ of agents in state $n = 0$ at time t who have not changed state since time t'. In the mean-field approximation, and treating time and numbers as continuous variables, these equations are given by

$$\begin{aligned} \frac{\partial N_0(t,t')}{\partial t} &= -2\frac{N_0(t,t')}{N}p_0(t,t') + \pi_{10}(t)\delta_{tt'}, \\ \frac{\partial N_1(t,t')}{\partial t} &= -2\frac{N_1(t,t')}{N}p_1(t,t') + \pi_{01}(t)\delta_{tt'}, \end{aligned} \tag{5}$$

where the transition rates $\pi_{n,m}(t)$ denote the average number of agents switching their states from n to m ($n \to m$) at time t. If the agents make their decisions according to the reinforcement dynamics described by the probabilities $p_n(t,t')$ given by Eq. (4), the dynamic equations (5) have a solution of the form

$$\begin{aligned} N_0(t,t') &= \pi_{10}(t')\left(1+\frac{t-t'}{N}\right)^{-2b_0}, \\ N_1(t,t') &= \pi_{01}(t')\left(1+\frac{t-t'}{N}\right)^{-2b_1}. \end{aligned} \tag{6}$$

Since the total number of isolated agents who change their state at time t is equal to $\pi_{01}(t)$ and the total number of interacting agents who change their state is equal to $\pi_{10}(t)$, it follows that $\pi_{10}(t)$ and $\pi_{01}(t)$ are given in terms of $N_0(t,t')$ and $N_1(t,t')$ by the relations

$$\begin{aligned} \pi_{10}(t) &= \frac{2}{N}\sum_{t'=1}^{t} p_1(t,t')N_1(t,t'), \\ \pi_{01}(t) &= \frac{2}{N}\sum_{t'=1}^{t} p_0(t,t')N_0(t,t'). \end{aligned} \tag{7}$$

To solve the coupled set of equations (6) and (7), we assume self-consistently that $\pi_{10}(t)$ and $\pi_{01}(t)$ are either constant or decaying in time as power laws. Therefore, we assume

$$\begin{aligned} \pi_{10}(t) &= \tilde{\pi}_{10}\left(\frac{t}{N}\right)^{-\alpha_0}, \\ \pi_{01}(t) &= \tilde{\pi}_{01}\left(\frac{t}{N}\right)^{-\alpha_1}. \end{aligned} \tag{8}$$

To check the self-consistent assumption Eq. (8), we insert it in Eqs. (6) and (7) and compute the values of the parameters α_0, α_1, $\tilde{\pi}_{10}$ and $\tilde{\pi}_{01}$ that determine the solution in the asymptotic limit $t \to \infty$. If $\alpha_0 = \alpha_1 = 0$, we obtain a stationary solution in which $\pi_{10}(t) = \tilde{\pi}_{10}$ and $\pi_{01}(t) = \tilde{\pi}_{01}$, are independent of time. On the contrary if $\alpha_0 > 0$ or $\alpha_1 > 0$, the system is non-stationary, with transition rates $\pi_{10}(t)$ and $\pi_{01}(t)$ decaying in time. The system dynamics slows down. In Appendix A, we give the details of this self-consistent calculation in the large N limit, which yields $\alpha_0 = \alpha_1 = \alpha$ and $\tilde{\pi}_{10} = \tilde{\pi}_{01} = \tilde{\pi}$, with

$$\begin{aligned} \alpha &= \max\left(0, 1-2b_1, 1-2b_0\right), \\ \tilde{\pi} &= \frac{\sin\left[2\pi\min\left(b0, b1\right)\right]}{\pi}[1-\delta(\alpha,0)] \\ &\quad + \frac{(2b_0-1)(2b_1-1)}{2(b_0+b_1-1)}\delta(\alpha,0). \end{aligned} \tag{9}$$

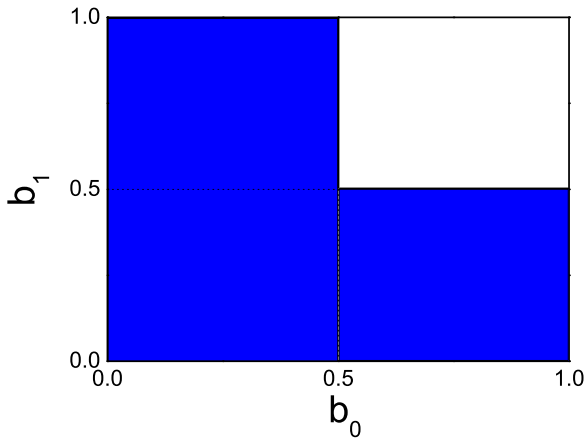

FIG. 4: (Color online) Phase diagram of the pairwise model. The white area indicates the stationary regime in which the transition rate is constant. The colored (gray) area indicates the non-stationary phase.

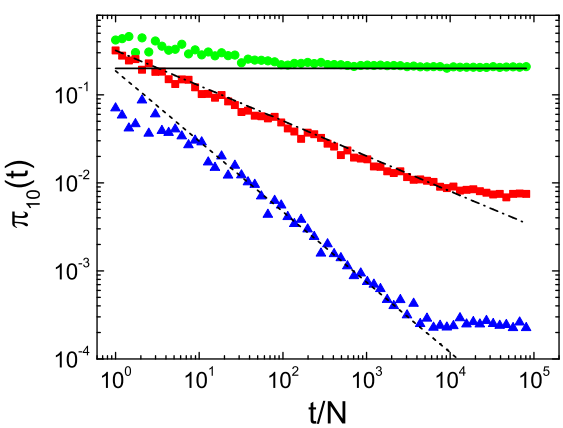

FIG. 5: (Color online) Evolution of the transition rate $\pi_{10}(t)$ in the different phase regions of the pairwise interaction model. The simulation is performed with $N = 1000$ agents for a number of time steps $T_{max} = N \times 10^5$, and averaged over 10 realizations. The simulations are performed in the stationary region with parameter values $b_0 = b_1 = 0.7$ (circles) and in the non-stationary region with parameter values $b_0 = 0.3$, $b_1 = 0.7$ (squares) and $b_0 = b_1 = 0.1$ (triangles). The lines indicate the analytical predictions Eqs. (8)-(9).

The analytically predicted dynamical behavior or the model can be summarized by the phase diagram depicted in Fig. 4 (that we discuss now in more detail), together with the numerical simulations of the stochastic model displayed in Fig. 5.

- *Stationary region ($b_0 > 0.5$ and $b_1 > 0.5$)* - In this region of the phase diagram, the self-consistent equation predicts $\alpha = 0$, so that a stationary state solution is expected, where $\pi_{01}(t) = \tilde{\pi}$ is given by Eq. (9). In this stationary state the number of isolated agents and the number or interacting agents are constant on average, but the dynamics is not frozen, since $\tilde{\pi} > 0$: agents continuously form and leave pairs. The simulations shown in Fig. 5 for $b_0 = b_1 = 0.7$ confirm this analytical prediction.

- *Non stationary region ($b_0 < 0.5$ or $b_1 < 0.5$)* - In this region of the phase diagram, the self-consistent equation predicts a non-stationary solution with $\pi_{10}(t)$ and $\pi_{01}(t)$ decaying with t as a power-law of exponent $\alpha = \max(1 - 2b_0, 1 - 2b_1)$. Figure 5 shows such a decay for $b_0 = 0.3$, $b_1 = 0.7$ and for $b_0 = b_1 = 0.1$, which is however truncated by finite-size effects for t larger than $t_c(N) \propto N$. Therefore the system eventually becomes stationary with a very slow dynamics (very small transition rates $\pi_{10}(t)$ and $\pi_{01}(t)$).

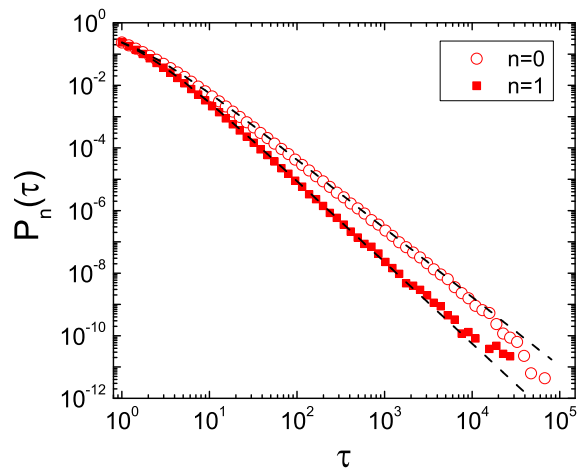

FIG. 6: (Color online) Probability distribution of the durations of contacts $P_1(\tau)$ and of the inter-contact durations $P_0(\tau)$ in the stationary region, for the pairwise model. The data is reported for a simulation with $N = 1000$ agents, run for $T_{max} = N \times 10^5$ elementary time steps, with parameter values $b_0 = 0.6$, $b_1 = 0.8$. The data is averaged over 10 realizations.

Empirical studies often focus on the statistics of contact durations between individuals, and of the time intervals between two contacts of a given individual. These quantities of interest can be computed in our model, respectively, as the probabilities $P_1(\tau)$ that an agent remains in a pair during a time $\tau = (t - t')/N$, and $P_0(\tau)$ that an agent remains isolated for a time interval $\tau = (t - t')/N$. These probabilities are determined by the numbers of agents in each state and the rates at which the agents change their state. The probability distributions of the time spent in each state, integrated between

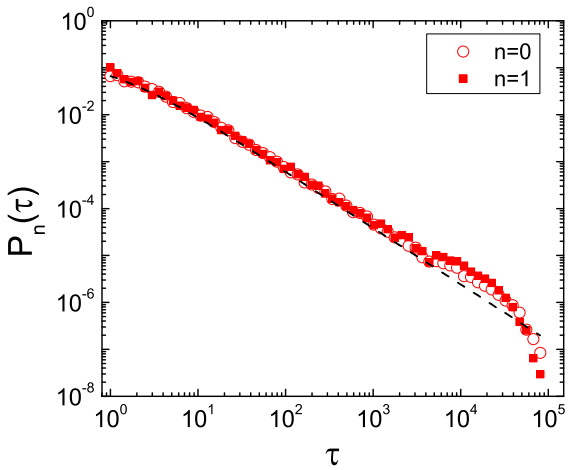

FIG. 7: (Color online) Probability distribution of the durations of contacts $P_1(\tau)$ and of the inter-contact durations $P_0(\tau)$ in the non-stationary region of the pairwise model, with $b_0 < 0.5$ and $b_1 < 0.5$. In this region we observe some deviations of the probabilities $P_1(\tau)$ and $P_0(\tau)$ from the power-law behavior for large durations. The data are reported for a simulation with $N = 1000$ agents run for $T_{max} = N \times 10^5$ elementary time steps, with parameter values $b_0 = b_1 = 0.1$. The data are averaged over 10 realizations.

the initial time and an arbitrary time t, are given by

$$P_n(\tau) \propto \int_{t'=0}^{t-N\tau} p_n(t'+N\tau,t')N_n(t'+N\tau,t')dt' \quad (10)$$

for $n = 0, 1$. Inserting the expression given by Eq. (6) for $N_n(t,t')$ and the definition of $p_n(t,t')$ given by Eq. (4) in Eq. (10), we obtain the power-law distributions

$$P_n(\tau) \propto (1+\tau)^{-2b_n-1} \quad (11)$$

for $n = 0, 1$. These analytical predictions are compared with numerical simulations in Fig. 6 for $b_0 = 0.6$, $b_1 = 0.8$ (stationary system) and in Fig. 7 for $b_0 = b_1 = 0.1$ (non-stationary π_{10} and π_{01}). Interestingly, even when the system is non-stationary, the distributions $P_n(\tau)$ remain stationary.

B. Formation of groups of any size

In this subsection we extend the solution obtained for the pairwise model to the general model with arbitrary value of the parameter λ, where groups of any size can be formed. Therefore the coordination number n_i of each agent i can take any value up to $N-1$. Extending the formalism used in the previous subsection, we denote by $N_n(t,t')$ the number of agents with coordination number $n = 0, 1, \ldots, N-1$ at time t, who have not changed state since time t'. In the mean field approximation, the evolution equations for $N_n(t,t')$ are given by

$$\begin{aligned}
\frac{\partial N_0(t,t')}{\partial t} &= -2\frac{N_0(t,t')}{N}p_0(t,t') - (1-\lambda)\epsilon(t) \\
&\quad \times \frac{N_0(t,t')}{N}p_0(t,t') + \sum_{i\geq 1}\pi_{i,0}(t)\delta_{tt'}, \\
\frac{\partial N_1(t,t')}{\partial t} &= -2\frac{N_1(t,t')}{N}p_1(t,t') \\
&\quad +[\pi_{0,1}(t)+\pi_{2,1}(t)]\delta_{tt'}, \\
\frac{\partial N_n(t,t')}{\partial t} &= -(n+1)\frac{N_n(t,t')}{N}p_1(t,t') \\
&\quad +[\pi_{n-1,i}(t)+\pi_{n+1,n}(t)+\pi_{0,n}(t)]\delta_{tt'}, \quad n \geq 2. \quad (12)
\end{aligned}$$

In these equations, the parameter $\epsilon(t)$ indicates the rate at which isolated nodes are introduced by another agent in already existing groups of interacting agents. Moreover, $\pi_{mn}(t)$ indicates the transition rate at which agents change coordination number from m to n (i.e. $m \to n$) at time t. In the mean-field approximation the value of $\epsilon(t)$ can be expressed in terms of $N_n(t,t')$ as

$$\epsilon(t) = \frac{\sum_{n\geq 1}\sum_{t'=1}^{t} N_n(t,t')p_1(t,t')}{\sum_{t'=1}^{t} N_0(t,t')p_0(t,t')}. \quad (13)$$

In the case of reinforcement dynamics described by the probabilities $p_n(t,t')$ given by Eq. (4), and assuming that asymptotically in time $\epsilon(t)$ converges to a time-independent variable, that is, $\lim_{t\to\infty}\epsilon(t) = \hat{\epsilon}$, the solution to the rate equations (12) in the large time limit is given by

$$\begin{aligned}
N_0(t,t') &= N_0(t',t')\left(1+\frac{t-t'}{N}\right)^{-b_0[2+(1-\lambda)\hat{\epsilon}]}, \\
N_1(t,t') &= N_1(t',t')\left(1+\frac{t-t'}{N}\right)^{-2b_1}, \quad (14) \\
N_n(t,t') &= N_n(t',t')\left(1+\frac{t-t'}{N}\right)^{-(n+1)b_1} \quad \text{for } n \geq 2,
\end{aligned}$$

with

$$\begin{aligned}
N_0(t',t') &= \sum_{n\geq 1}\pi_{n,0}(t'), \\
N_1(t',t') &= \pi_{0,1}(t') + \pi_{2,1}(t'), \quad (15) \\
N_n(t',t') &= \pi_{n-1,n}(t') + \pi_{n+1,n}(t') + \pi_{0,n}(t') \quad \text{for } n \geq 2.
\end{aligned}$$

The transition rates $\pi_{m,n}(t)$ can be determined in terms of $N_n(t,t')$ as shown in the Appendix B. In order to solve the equations we make the further assumption that the transition rates $\pi_{mn}(t)$ are either constant or decaying with time according to a power law, that is.

$$\pi_{m,n}(t) = \tilde{\pi}_{m,n}\left(\frac{t}{N}\right)^{-\alpha_{m,n}}. \quad (16)$$

Self-consistent calculations (see Appendix B) determine the value of the quantities $\hat{\epsilon}$, α_{mn}, and $\tilde{\pi}_{mn}$. For $\lambda > 0.5$

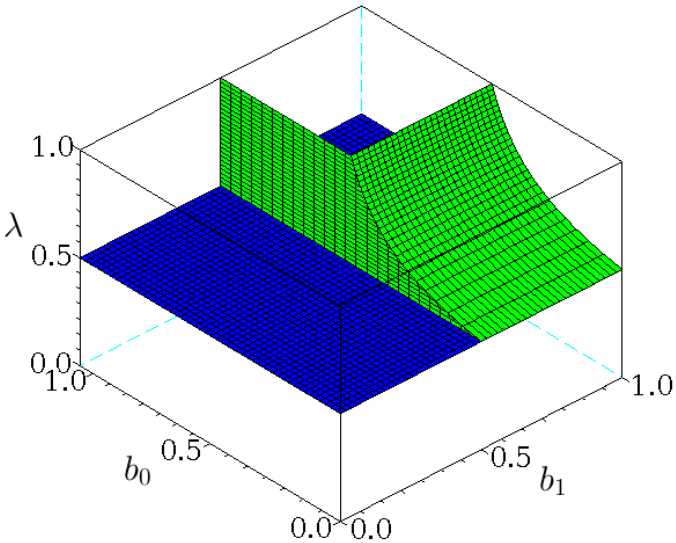

FIG. 8: (Color online) Phase diagram of the general model with formation of groups of arbitrary size. The region behind the green surface corresponds to the stationary phase [i.e., region (I), with $\lambda > 0.5$, $b_1 > 0.5$ and $b_0 > \frac{2\lambda-1}{3\lambda-1}$]. The region in front of the green surface and above the blue one [region (II)] corresponds to a non-stationary system with decaying transition rates. Strong finite size effects with a temporary formation of a large cluster are observed in the region below the blue surface [i.e., region (III) with $\lambda < 0.5$].

the self-consistent assumption Eq. (16) is valid and we find, as in the case of pairwise interactions, that $\alpha_{m,n} = \alpha$ $\forall(m,n)$, with

$$\alpha = \max\left(0, 1 - b_0\frac{3\lambda-1}{2\lambda-1}, 1-2b_1\right). \qquad (17)$$

This solution generalizes the case of the pairwise model, which is recovered by setting $\lambda = 1$. For $\lambda \leq 0.5$ the self-consistent assumption breaks down and we will resort to numerical simulations.

The probability distributions of the time spent in each state, integrated between the initial time and an arbitrary time t, are given by

$$P_n(\tau) \propto \int_{t'=0}^{t-N\tau} p_n(t'+N\tau, t')N_n(t'+N\tau, t')dt'. \quad (18)$$

Inserting the expression given by Eq. (14) for $N_n(t,t')$ and the definition of $p_n(t,t')$ given by Eq. (4) in Eq. (18), we obtain the power-law distributions

$$\begin{aligned} P_0(\tau) &\propto (1+\tau)^{-b_0[2+(1-\lambda\hat{\epsilon})]-1} \\ P_n(\tau) &\propto (1+\tau)^{-(n+1)b_1-1} \quad \text{for } n \geq 1. \end{aligned} \qquad (19)$$

The phase diagram of the model is summarized in Fig. 8. We can distinguish between three phases.

- *Region (I) -The stationary region:* $b_1 > 0.5$, $b_0 > (2\lambda-1)/(3\lambda-1)$ *and* $\lambda > 0.5$- In this region, the self-consistent solution yields $\alpha = 0$: the transition rates $\pi_{mn}(t)$ converge rapidly to a constant value (see Fig. 9 for a comparison between numerics and analytics for $\pi_{10}(t)$) and the system reaches a stationary state. In Fig. 10 we compare the analytical solution given by Eqs. (19) with the numerical simulations in the stability region, finding perfect agreement. As predicted by Eqs. (19), $P_n(\tau)$ decays faster as n increases: larger groups are less stable than smaller ones, as found in the empirical

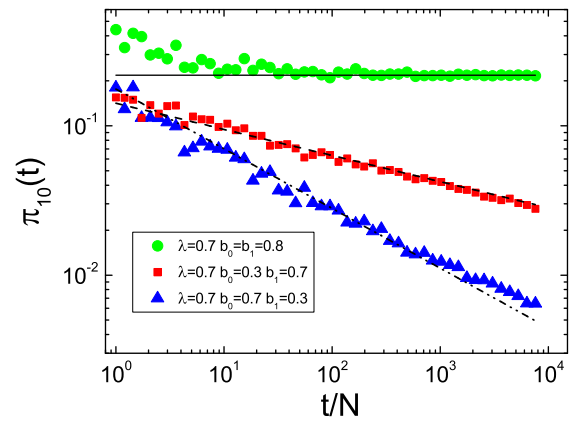

FIG. 9: (Color online) Transition rate $\pi_{10}(t)$ for the model in the presence of groups of any size, for different parameters λ, b_0, b_1 corresponding to the different regions of the phase diagram. The straight lines correspond to the analytical predictions. The simulation is performed with $N = 1000$ agents for a number of time steps $T_{max} = N \times 10^4$. The data are averaged over 10 realizations.

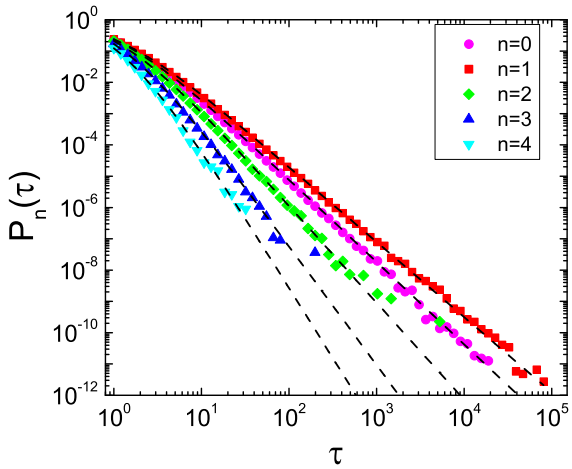

FIG. 10: (Color online) Distribution $P_n(\tau)$ of durations of groups of size $n+1$ in the stationary region. The simulation is performed with $N = 1000$ agents for a number of time steps $T_{max} = N \times 10^5$. The parameter used are $b_0 = b_1 = 0.7$, $\lambda = 0.8$. The data are averaged over 10 realizations. The dashed lines correspond to the analytical predictions Eqs. (19).

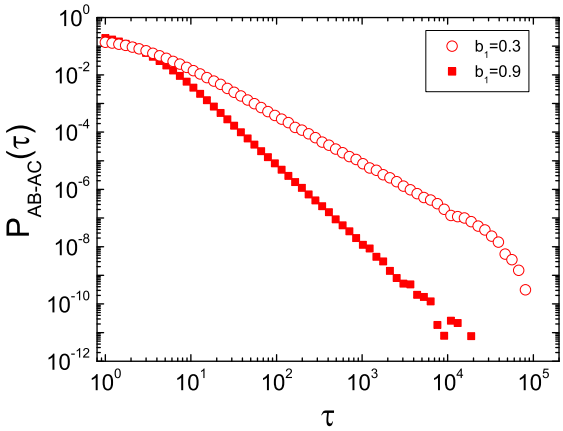

FIG. 11: (Color online) Distribution of time intervals between successive contacts of an individual, for $\lambda = 0.8$, $b_0 = 0.7$ and $b_1 = 0.3$ and 0.9. The simulation is performed with $N = 10^4$ for a number of time steps $T_{max} = N \times 10^5$. The data are averaged over 10 realizations.

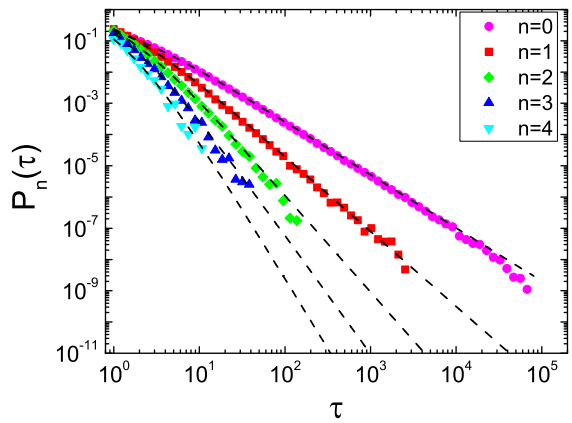

FIG. 13: (Color online) Distribution $P_n(\tau)$ of durations of groups of size $n+1$ in the non stationary region [region (II)]. The simulation is performed with $N = 1000$ agents for a number of time steps $T_{max} = N \times 10^5$. The parameter used are $b_0 = 0.3$ and $b_1 = 0.7$, $\lambda = 0.8$. The data are averaged over 10 realizations. The dashed lines correspond to the analytical predictions Eqs. (19).

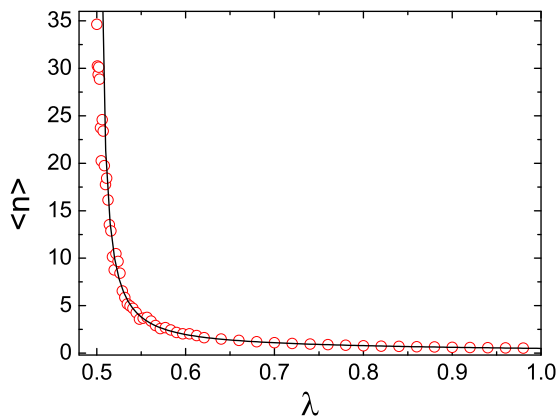

FIG. 12: (Color online) Average coordination number $\langle n \rangle$ vs λ for $b_0 = b_1 = 0.7$. The simulation is performed with $N = 2000$ agents for a number of time steps $T_{max} = N \times 10^3$. $\langle n \rangle$ is computed in the final state over 30 realizations. The solid line indicates the theoretical prediction given by Eq. (20).

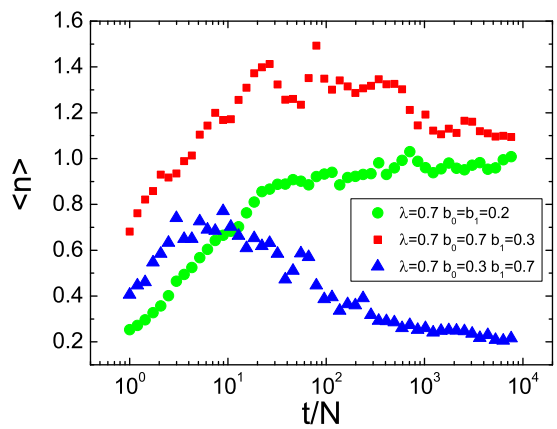

FIG. 14: (Color online) Average coordination number $\langle n \rangle$ as a function of time in the Region II of the phase diagram for different values of the parameters λ, b_0 and b_1. The data is in very good agreement with the theoretical expectations given by Eqs. (21) − (22). The simulations are performed with $N = 1000$ agents for a number of time steps $T_{max} = N \times 10^4$. The data are averaged over 10 realizations.

data sets. Figure 11 displays the distribution $P_{AB-AC}(\tau)$ of time intervals between the start of two consecutive contacts of a given individual, which is as well stationary and displays a power-law behavior.

The average coordination number $\langle n \rangle$ is given by

$$\langle n \rangle = \frac{\pi_{10}}{2\lambda} \sum_{n \geq 1} \frac{n(n+1)}{(n+1)b_1 - 1} \left(\frac{1-\lambda}{\lambda} \right)^{n-1}, \qquad (20)$$

where the detailed calculation and the value of $\pi_{10}(t)$ are given in Appendix B. This expression diverges as $\lambda \to 0.5$. In Fig. 12 we show the perfect agreement between the result of numerical simulations of $\langle n \rangle$ and the theoretical prediction.

- *Region (II) -Non-stationary region:* $b_1 < 0.5$ *or* $b_0 < (2\lambda - 1)/(3\lambda - 1)$, *and* $\lambda > 0.5$ - The dynamics in this region is non-stationary and the transition rate is decaying with time as a power-law, as shown in Fig. 9 where we report $\pi_{10}(t)$ as a function of t. Nevertheless, the distributions of lifetimes of groups of various sizes $P_n(\tau)$, and of inter-contact times $P_{AB-AC}(\tau)$, re-

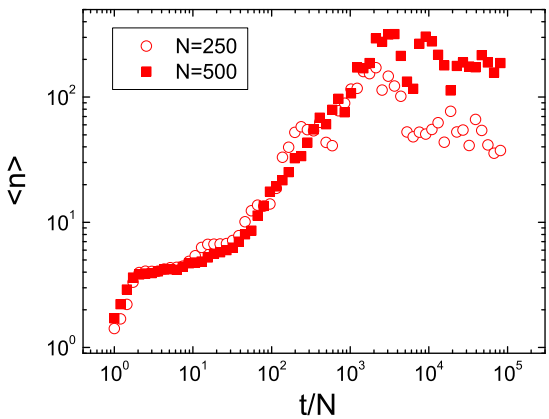

FIG. 15: (Color online) Average coordination number $\langle n \rangle$ for $\lambda = 0.2$, $b_0 = b_1 = 0.7$. The simulations of a single realization are performed with $N = 250$ and $N = 500$ agents, respectively, for a number of time steps $T_{max} = N \times 10^5$.

main stationary. These distributions are shown in Figs. 13 and 11. In this region, the average coordination number in the limit $t/N \gg 1$ remains small, even as $\lambda \to 0.5$. In particular from the mean-field solution of the dynamics (see appendix B) the theoretical solution of the model predicts that, for $\lambda > 0.5$ and $t \to \infty$

$$\langle n \rangle = 1 \quad \text{for } \alpha = 1 - 2b_1, \tag{21}$$

and

$$\langle n \rangle = 0 \quad \text{for } \alpha = 1 - b_0 \frac{3\lambda - 1}{2\lambda - 1}. \tag{22}$$

Figure 14 shows the agreement of this predicted behavior with simulation results for several values of b_0 and b_1 and $\lambda = 0.7$. In this region, as $\lambda \to 0.5$ with fixed b_0 and b_1, we have $\alpha = 1 - 2b_1$ and $\langle n \rangle \to 1$. Therefore no diverging behavior is observed.

- *Region (III) Strong dependence on the number of agents N and non-stationary dynamics:* $\lambda < 0.5$ - In this region the self-consistent assumption given by Eq. (16) breaks down, and we find numerically that the average coordination number $\langle n \rangle$ strongly depends on the number of agents N and on time. In order to give a typical example of the corresponding dynamical behavior, Fig. 15 displays $\langle n \rangle$ as a function of time for two single realizations of the model corresponding to two different values of N. Interestingly, the distributions of lifetimes of groups of various sizes $P_n(\tau)$ remain stationary even in this parameter region (not shown).

C. Aggregated networks

In the previous paragraphs, we have shown how our modeling framework produces dynamical properties of the interactions between agents that yield broad distributions of contact and inter-contact times. In order to understand the structure of the resulting interaction networks at coarser temporal resolutions, it is as well interesting to investigate the properties of the aggregated networks, constructed as in Sec. II. Given a starting time t_0 and a temporal window ΔT, the nodes of these networks are the agents, and a link is drawn between two agents whenever they have been in contact between t_0 and $t_0 + \Delta T$, with a link weight given by the total time during which they have interacted in $[t_0, t_0 + \Delta T]$. As in Sec. II, the degree k_i of an agent i is given by the number of distinct agents with whom i has been in contact in $[t_0, t_0 + \Delta T]$, while its strength s_i is the sum of the interaction times with other agents, and the participation ratio $Y_2(i)$ quantifies the heterogeneity of the times spent by i with these other agents.

As an exhaustive exploration of the aggregated networks and of how their properties depend on the model's parameter would be tedious, we simply report in Fig. 16 the properties of aggregated networks for increasing window lengths ΔT and for two sets of parameters. Some properties are qualitatively similar to the empirically observed networks. In particular, the degree distributions are peaked around an average value that increases with ΔT. As time passes, each agent encounters more and more distinct other agents, and the distribution $P(k)$ globally shifts towards larger degrees. The links weights distributions are broad, and extend to larger values as ΔT increases. Some other properties seem to depend strongly on the model's parameters. In particular, the average strength of nodes of degree k, and the average participation ratio of nodes of degree k, can have shapes rather different from the empirical ones. Moreover, the time window lengths ΔT on which the aggregated network remains sparse are rather restricted.

IV. EXTENSIONS OF THE MODEL

A. Heterogeneous model

In the previous section, we have assumed that all the agents have the same tendency to form a group or to leave a group, that is, the probabilities p_n do not depend on the agent who performs a status update. Real social systems display however additional complexity since the social behavior of individuals may vary significantly across the population. A natural extension of the model presented above consists therefore of making the probabilities p_n dependent on the agent who is updating his/her state. To this aim, we assign to each agent i a parameter η_i that characterizes his/her propensity to form social interactions. In real networks this propensity will depend on the features of the agents [65]. In the model we assume that this propensity, that we call "sociability", is a quenched random variable, which is assigned to each agent at the start of the dynamical evolution and re-

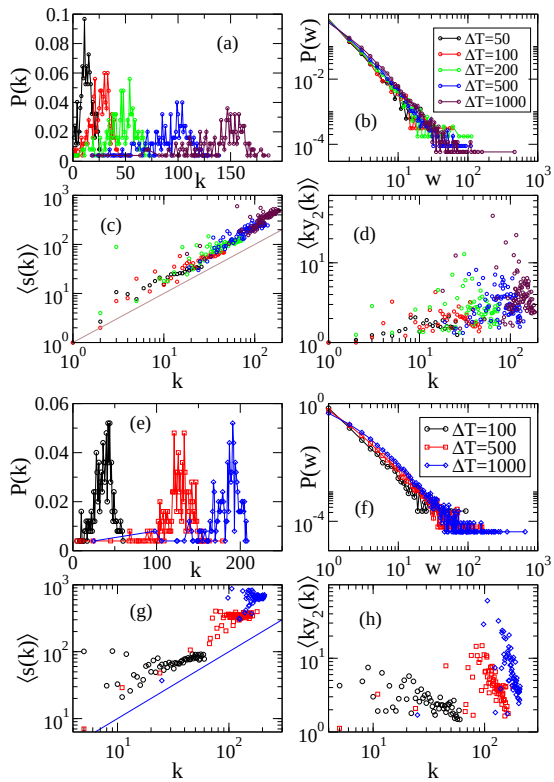

FIG. 16: (Color online) Aggregated networks' characteristics for the model with constant number of agents ($N = 250$), for time windows ΔT of increasing lengths, and two sets of parameters: $(b_0, b_1, \lambda) = (0.55, 0.8, 0.9)$ (a, b, c, d) and $(0.7, 0.7, 0.8)$ (e, f, g, h).

mains constant, and we assume for simplicity that it is uniformly distributed in $[0, 1]$. In this modified model, the probability $p_n^i(t, t')$ that an agent i with coordination number n since time t' changes his/her coordination number at time t is given by

$$\begin{aligned} p_0^i(t,t') &= \frac{\eta_i}{1+(t-t')/N}, \\ p_n^i(t,t') &= \frac{1-\eta_i}{1+(t-t')/N}, \text{ for } n \geq 1. \end{aligned} \tag{23}$$

In this setup, the parameters (b_0, b_1), which did not depend on i in Eq. (4), are replaced for each agent i by the values $(\eta_i, 1-\eta_i)$: a large η_i corresponds to an agent who prefers not to be isolated.

The agents' heterogeneity adds a significant amount of complexity to the problem, and we have reached an analytical solution of the evolution equations only in the case of pairwise interactions ($\lambda = 1$). The general case can be studied through numerical simulations as we discuss at the end of this section.

Let us denote by $N_0(t, t', \eta)$ the number of isolated agents with parameter $\eta_i \in [\eta, \eta + \Delta\eta]$ who have not changed their state since time t'. Similarly, we indicate by $N_1(t, t', \eta, \eta')$ the number of agents in a pair joining two agents i and j with $\eta_i \in [\eta, \eta+\Delta\eta], \eta_j \in [\eta', \eta'+\Delta\eta]$, who have been interacting since time t'. The mean-field equations for the model are then given by

$$\begin{aligned} \frac{\partial N_0(t,t',\eta)}{\partial t} &= -2\frac{N_0(t,t',\eta)}{N} p_0(t,t',\eta) \\ &\quad + \pi_{10}^{\eta}(t)\delta_{tt'}, \\ \frac{\partial N_1(t,t',\eta,\eta')}{\partial t} &= -\frac{N_1(t,t',\eta,\eta')}{N}[p_1(t,t',\eta) + p_1(t,t',\eta')] \\ &\quad + \pi_{01}^{\eta\eta'}(t)\delta_{tt'}. \end{aligned} \tag{24}$$

With the expression for $p_n(t, t', \eta)$ given by Eqs.(23) we find

$$\begin{aligned} N_0(t,t',\eta) &= \pi_{10}^{\eta}(t')\Big(1+\frac{t-t'}{N}\Big)^{-2\eta}, \\ N_1(t,t',\eta,\eta') &= \pi_{01}^{\eta\eta'}(t')\Big(1+\frac{t-t'}{N}\Big)^{-2+\eta+\eta'}. \end{aligned} \tag{25}$$

The transition rate π_{10}^{η} gives the rate at which agents with $\eta_i \in [\eta, \eta+\Delta\eta]$ become isolated, and $\pi_{01}^{\eta\eta'}$ is the rate at which pairs ij with $\eta_i \in [\eta, \eta+\Delta\eta], \eta_j \in [\eta', \eta'+\Delta\eta]$ are formed. These rates can be expressed as a function of $N_0(t, t', \eta)$ and $N_1(t, t', \eta, \eta')$ according to

$$\pi_{10}^{\eta}(t) = \sum_{t',\eta'} \frac{N_1(t,t',\eta,\eta')}{N}[p_1(t,t',\eta) + p_1(t,t',\eta')], \tag{26}$$

$$\pi_{01}^{\eta\eta'}(t) = 2\sum_{t',t''} \frac{N_0(t,t',\eta)N_0(t,t'',\eta')}{C(t)N} p_0(t,t',\eta)p_0(t,t'',\eta'),$$

where $C(t)$ is a normalization factor given by

$$C(t) = \sum_{t'=1}^{t}\sum_{\eta} N_0(t,t',\eta)p_0(t,t',\eta). \tag{27}$$

To solve this problem with the same strategy used for the model without heterogeneity we make the self-consistent assumption that the transition rates are either constant or decaying as a power law with time:

$$\pi_{10}^{\eta}(t) = \Delta\eta\tilde{\pi}_{10}^{\eta}\Big(\frac{t}{N}\Big)^{-\alpha(\eta)}, \tag{28}$$

$$\pi_{01}^{\eta\eta'}(t) = \Delta\eta\Delta\eta'\tilde{\pi}_{01}^{\eta\eta'}\Big(\frac{t}{N}\Big)^{-\alpha(\eta,\eta')}. \tag{29}$$

In appendix C we give the details of the self-consistent calculation, which leads to the analytical prediction

$$\begin{aligned} \alpha(\eta) &= \max\left(1-2\eta, \eta-\frac{1}{2}\right), \\ \alpha(\eta,\eta') &= \alpha(\eta)+\alpha(\eta'), \end{aligned} \tag{30}$$

and the value of $\tilde{\pi}_{10}^{\eta}$ is given by

$$\tilde{\pi}_{10}^{\eta} = \begin{cases} \frac{\rho(\eta)}{B(1-2\eta,2\eta)} & \eta \leq \frac{1}{2} \\ \frac{\rho(\eta)}{B(\eta-\frac{1}{2},1)} & \eta \geq \frac{1}{2} \end{cases}. \tag{31}$$

In order to check the validity of our mean-field calculation, we study the probability distribution $P_0(\tau)$ of the

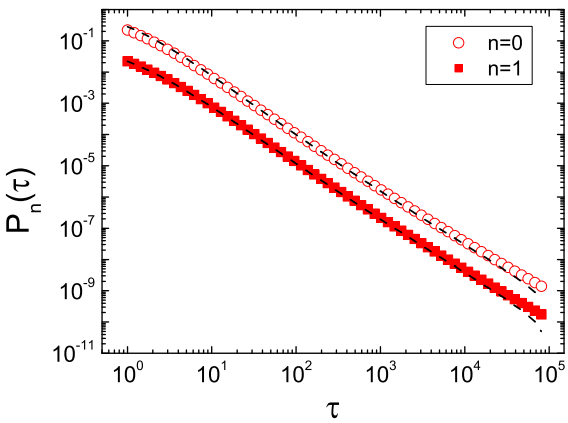

FIG. 17: (Color online) Distributions of times spent in state 0 and 1 for the heterogeneous model. The simulation is performed with $N = 10^4$ for a number of time steps $T_{max} = N \times 10^5$. The data are averaged over 10 realizations. The symbols represent the simulation results (circles for $n = 0$ and squares for $n = 1$). The dashed lines represent our analytical prediction. In order to improve the readability of the figure we have multiplied $P_1(\tau)$ by a factor of 10^{-1}.

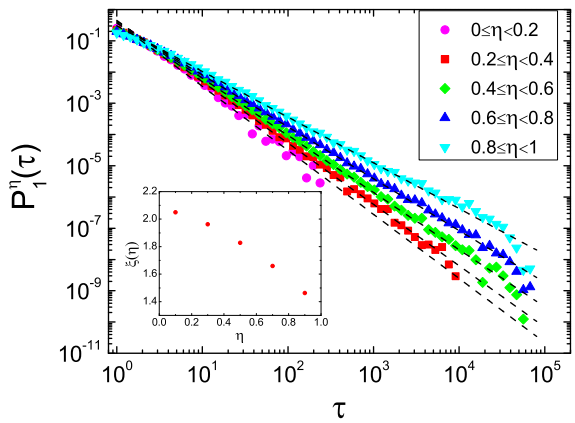

FIG. 18: (Color online) Distribution $P_1^\eta(\tau)$ of contact durations of individuals with sociability η in the pairwise heterogeneous model. The simulations are performed with $N = 1000$ agents and $T_{max} = N \times 10^5$ time steps. The data are averaged over 10 realizations. The data decays as a power-law $P_1^\eta(\tau) \propto \tau^{-\xi(\eta)}$, and we report the exponents $\xi(\eta)$ as a function of η in the inset.

durations of inter-contact periods and the distribution $P_1(\tau)$ of the durations of pairwise contacts, which are given, when averaged for a total simulation time T_{max}, by

$$\begin{aligned} P_0(\tau) &\propto \int_0^{T_{max}-N\tau} dt \int_0^1 d\eta \pi_{10}^{\eta}(t)\eta(1+\tau)^{-2\eta-1}, \\ P_1(\tau) &\propto \int_0^{T_{max}-N\tau} dt \int_0^1 d\eta \int_0^1 d\eta' \pi_{01}^{\eta\eta'} \\ &\times (2-\eta-\eta')(1+\tau)^{\eta+\eta'-3}, \end{aligned} \tag{32}$$

where $\rho(\eta)$ is the probability distribution of η. In Fig. 17 we compare the probabilities of intercontact time $P_0(\tau)$ and contact time $P_1(\tau)$ averaged over the full population together with the numerical solution of the stochastic model, showing a perfect agreement. In Fig. 18 moreover, we show the distributions $P_1^\eta(\tau)$ of the contact durations of agents with $\eta_i \in (\eta, \eta + \Delta\eta)$. Power-law behaviors are obtained even at fixed sociability, and the broadness of the contact duration distribution of an agent increases with the"sociability" of the agent under consideration.

As previously mentioned, the model can be extended by allowing the formation of large groups, by setting $\lambda < 1$. The results of numerical simulations performed for a particular value of λ are shown in Fig.19. Power law distributions of the lifetime of groups are again found and, as in the basic model without heterogeneity of the agents, larger groups are more unstable than smaller groups, as $P_n(\tau)$ decays faster if the coordination number n is larger. As the parameter $\lambda \to 0.5$ there is a phase transition and the average coordination number

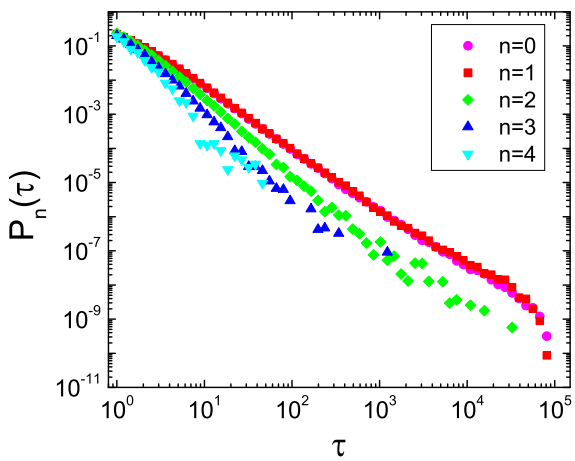

FIG. 19: (Color online) Distribution $P_n(\tau)$ of the durations of groups of size $n+1$ in the heterogeneous model with formation of groups of any size. The data are shown for simulations of $N = 1000$ agents performed over $T_{max} = N \times 10^5$ time steps and $\lambda = 0.8$, averaged over 10 realizations.

diverges. In Fig. 20 we show that $\langle n \rangle \propto (\lambda - 0.5)^{-\delta}$ with $\delta = 1$ within the statistical fitting error, similarly to what happens in the homogeneous case. Overall, the main features of the model are therefore robust with respect to the introduction of heterogeneity in the agents' individual behavior.

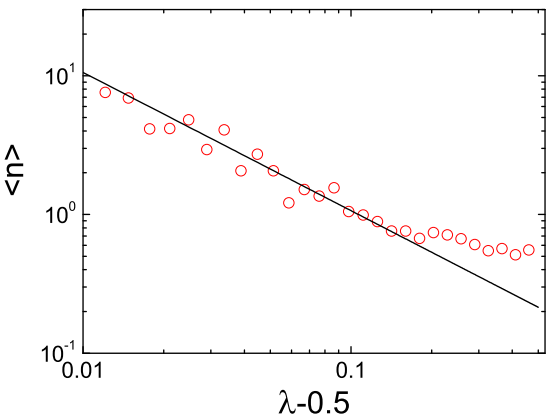

FIG. 20: (Color online) Average coordination number $\langle n \rangle$ of the agents in the heterogeneous case as a function of λ. The solid line indicates the best fit with $\langle n \rangle \propto (\lambda - 0.5)^{-\delta}$ with $\delta = 0.996$ in agreement with the exponent -1 within the statistical uncertainty. The data correspond to simulations of $N = 500$ agents performed over $T_{max} = N \times 10^3$ time steps. The data are averaged over 10 realizations.

B. Model with variable number of agents

When measurements on the proximity or face-to-face contacts of individuals are performed, the number of agents present on the premises typically fluctuates. Moreover, the activity of the agents fluctuates as well, because for instance of day/night patterns or periods of coffee/lunch breaks in a conference. This is in particular the case in the data sets described in Sec. II.

In this subsection, we use these simple remarks to put forward an extension of our model of interacting agents that mimics in a more realistic way the data gathering process, and that can be used to produce more realistic artificial data sets. The main point is to consider, instead of a population with a fixed number of agents N, an *open* system leaving the possibility for agents to leave or enter it. To this aim, we simply introduce for each agent the possibility to be in a state called "absent". In this "absent" state, agents may be isolated or not, but the measurement infrastructure is not able to know it. Statistics on the time spent by an agent in a given state, or of the duration of contact and inter-contact times, are thus obtained by considering only "present" agents. Overall, the model's dynamics is as described in Sec. III, with parameters b_0, b_1, λ, with the addition that at each time step, "absent" agents can enter the system, or agents can leave the system and become "absent". Various rules could be thought of, but for the sake of simplicity we consider that an agent i entering the system become isolated ($n_i \to 0$), and that agents of any state n can leave the system (one could also allow "absent" agents to directly enter a group, restrict the possibility to leave to isolated agents, etc...). In this framework, two important points have to be considered

- the agents who leave the system can re-enter it: the global number of agents (absent and present) is constant. We will see that a set-up in which an agent who has left the premises cannot re-enter the system leads to an interesting modification of the system's properties.

- The decision for an agent to leave or enter the system can simply be given by constant probabilities. In the resulting dynamics, the number of agents who are present is simply fluctuating around a constant value that depends on these probabilities. Another possibility consists of *fixing* at each time step the number of present agents $N(t)$. The imposed $N(t)$ can be a given function of time such as for instance a periodic signal (possibly with superimposed stochastic noise) to mimic circadian rhythms. To mimic a realistic process, $N(t)$ can also be given by *an empirical time series from a real data set*, as we now consider (note that the total number N of agents has to be at least equal to $\max_t N(t)$).

In order to impose the number of agents present in the system at each time, we first construct the empirical time series. To this aim, various time steps can be used. We consider the natural temporal resolution obtained in SocioPatterns deployments, that is, $20s$, but other resolutions could be chosen as well. Considering that agents act independently, and make choices in a simultaneous way, we then identify the empirical time step with one attempted status update per present agent. After each series of $N(t)$ attempted status updates (according to the rules described in Sec. III), we check if the number of present agents is smaller or larger than the desired $N(t+1)$. If it is larger, we remove at random agents (i.e., put them in the "absent" state) in order to match the desired $N(t+1)$. If it is smaller, absent agents are introduced into the system, and put into the isolated state. The system evolves in this way for a number of steps equal to the number of time steps of the empirical data set, and we monitor the time evolution of the number of contacts between agents, and of agents in each state, as a function of time. We also note that this model can in fact be simulated for arbitrarily long times by simply repeating the imposed time series $N(t)$.

Figure 21 compares the resulting activity patterns with the empirically monitored activity, for two values of the parameter set (b_0, b_1, λ), when $N(t)$ is taken from the data gathered at the scientific conference ESWC described in Sec. II. We note that, although only $N(t)$ is imposed to be exactly the same in the model and in the data, it is possible to tune the parameters so that the other measures (number of isolated nodes, of links, of triangles) are simultaneously similar to real data. Figure 21 illustrates that the system's dynamics is highly non-stationary. As shown in [22], empirically observed

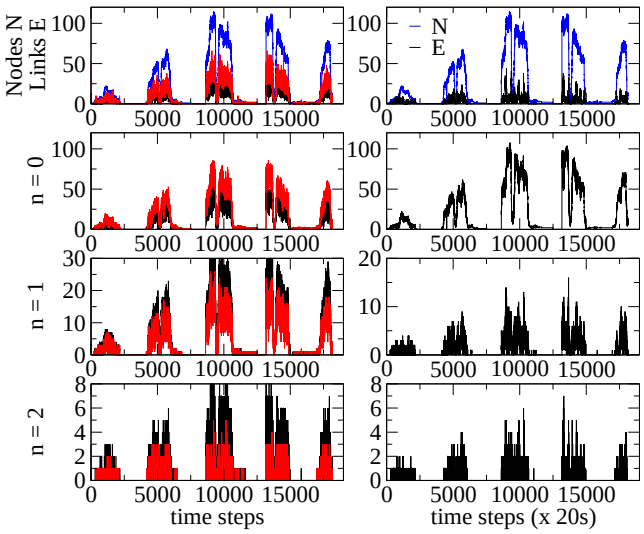

FIG. 21: (Color online) Timelines of the number of (from top to bottom): nodes and links in the instantaneous network (top), isolated nodes, groups of two nodes, groups of three nodes. The left column corresponds to the model with $N(t)$ imposed from the ESWC dataset and two sets of parameters, namely $(b_0, b_1, \lambda) = (0.55, 0.8, 0.9)$ (black curves) and $(0.7, 0.7, 0.8)$ (red curves), the right column to the real ESWC data set.

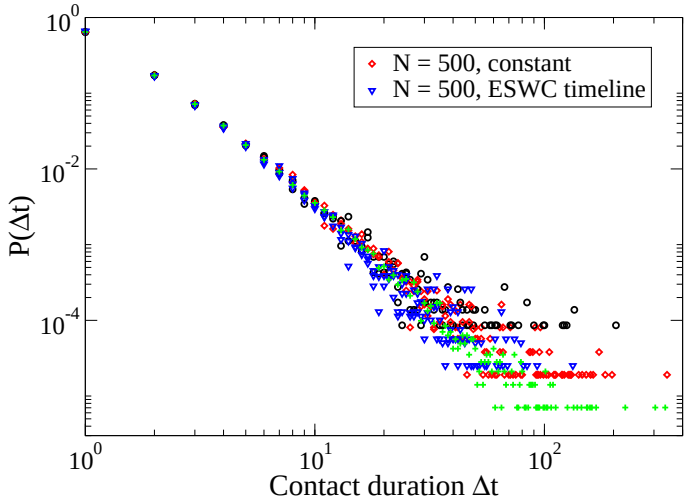

FIG. 22: (Color online) Distributions of contact durations, for $N(t)$ given by the ESWC data set, and for the model with constant N, for various time windows. Parameter values are $(b_0, b_1, \lambda) = (0.55, 0.8, 0.9)$.

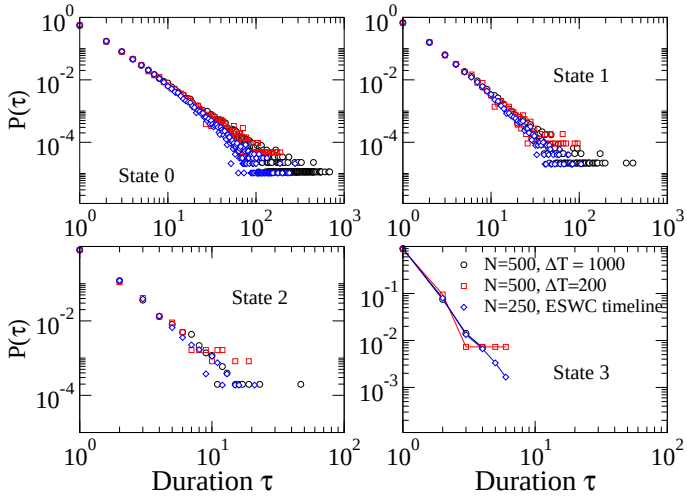

FIG. 23: (Color online) Distributions of the time spent in various states, for $N(t)$ given by the ESWC data set, and for the model with constant N. Parameter values are $(b_0, b_1, \lambda) = (0.55, 0.8, 0.9)$.

distributions of contact durations are however stationary. We check that this is also the case in our model by measuring the distributions of contact durations and of time spent by agents in each state in various time windows of different lengths, during which $N(t)$ strongly varies. As shown in Figs. 22 and 23, for a given parameter set (b_0, b_1, λ), these distributions are broad, as in the original model with constant number of agents, and do not depend strongly on the imposed $N(t)$, and can be superimposed from one time window to the next. As in the real data, the only differences come from different cutoffs stemming from different statistics in the different time windows.

As for the study of empirical data (see Sec II), we also construct and study the aggregated networks of contacts between agents, on different time windows. Figure 24, to be compared with Fig. 16 of Sec. III, illustrates the basic properties of these networks, which are very similar to the empirical ones shown in Fig. 3 of Sec. II. With respect to the case of constant N of Fig. 16, the degree distributions shift more slowly toward large degrees, and remain broader, so that the average strength of nodes of degree k, $[s(k)]$ and the average participation ratio of the nodes of degree k, $[ky_2(k)]$, keep more realistic shapes.

We finally illustrate the versatility of the modeling framework by considering the case in which an agent who leaves the network cannot re-enter it at future times. Such additional assumption can be considered to model environments in which there is a flux of individuals, such as a museum. As shown in [23], the daily aggregated network of the interactions between individuals is then not a small-world, as visitors entering the museum at very different times do not encounter each other. Figure 25 exemplifies how this kind of configurations can also be reproduced by our modeling framework, by simply imposing the empirically measured number of visitors present on the premises, $N(t)$, at each time step, and that an agent leaving the system does not re-enter it. The resulting network has an elongated shape whose topology is dictated by the timeline of the visits, exactly as in [23], and is strongly different from the case in which agents can leave and re-enter the system, as in a conference.

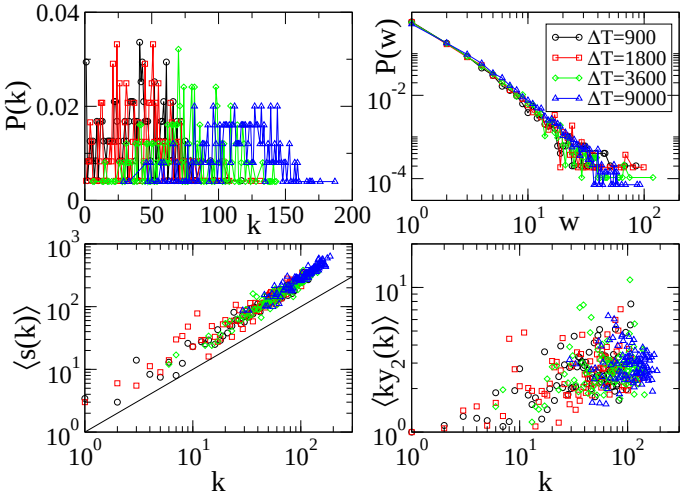

FIG. 24: (Color online) Aggregated networks' characteristics for $N(t)$ given by the ESWC time series, and $b_0 = b_1 = 0.7$, $\lambda = 0.8$, for various time window lengths. Top left: degree distributions; top right: links' weights distributions; Bottom left: average strength $\langle s(k)\rangle$ of nodes of degree k. Bottom right: average participation ratio $\langle ky_2(k)\rangle$ of nodes of degree k.

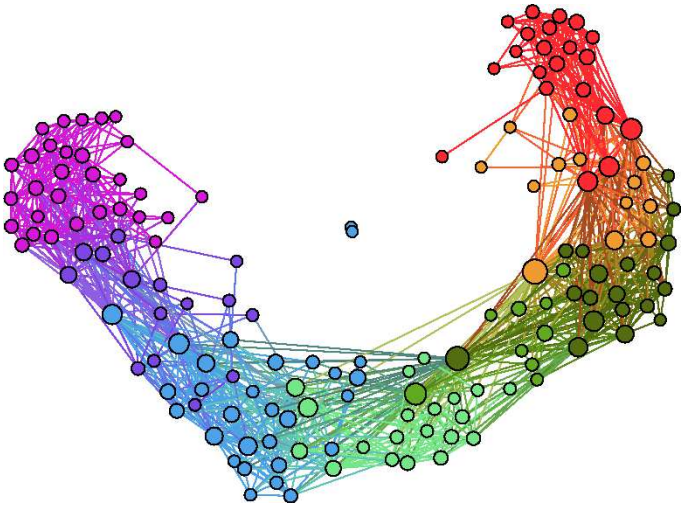

FIG. 25: (Color online) Example of aggregated network of 157 nodes obtained by imposing the timeline of the number of agents present at each time during a given day of data gathered during a SocioPatterns deployment at the Science gallery in Dublin [23]. Here $(b_0, b_1, \lambda) = (0.55, 0.8, 0.9)$, and an agent who leaves the network cannot re-enter it. The nodes are colored and positioned according to the entry time of the corresponding visitor, as in [23], from red (top right) to green (right) to blue (left) to violet (top left). The elongated shape of the network is similar to the one observed empirically in [23].

V. CONCLUSION

In this paper, we have studied with analytical and numerical means an agent-based model aimed at describing the dynamics of human interactions in social gatherings, as can be measured by recent technological wearable sensors that measure proximity patterns of individuals. This model is based on simple mechanisms in order to be easily implementable and extended through the introduction of more realistic and involved ingredients. The resulting distributions of contact durations and of inter-contact times can be either narrowly or broadly distributed. A detailed study of the latter case, obtained by a mechanism reminiscent of preferential attachment, reveals moreover interesting non-equilibrium transitions between stationary and non-stationary phases.

We have illustrated the versatility of the model by introducing two variations. In the first one, agents have a priori different propensities to interact, in order to mimic the heterogeneity in the social behaviors of individuals. We have shown that the phenomenology of the model is then conserved, with broad distributions of contact times, inter-contact times, and lifetimes of groups. In the second extension of the model, the population size can fluctuate. In this case, the population size can be taken as input of the model, as given by an empirical time series coming from a real-world data set. We have then shown that the model produces non-stationary dynamical networks whose features are close to the empirically observed ones.

The present modeling framework, and in particular its extension to a varying population, opens various perspectives. For instance, we have considered indistinguishable agents, who enter and leave the system at random times. It would however be possible to impose for each agent the *time intervals* in which he/she is present, which could be taken from real data. Such a procedure would produce artificial data sets that closely mimic the empirical timelines. Although this would happen at the cost of a rather large input of empirical information, it would produce data sets that retain the details of the presence properties of individual agents of the empirical data sets, but with tunable contact and inter-contact time distributions. Another interesting outcome of the model with varying population size is that it makes it possible, starting from an empirical data set that is by definition limited in time, to create a dynamical network on arbitrarily long timescales, with the same properties as the real one, by simply repeating the time-series $N(t)$ as many times as required. This corresponds to an interesting way of creating a non-stationary dynamical network, without having to repeat the *real* sequence of contacts. On each new repetition of $N(t)$, the model will generate a new sequence of contacts.

The modeling framework we have presented, and the above-mentioned perspectives, are particularly interesting in the perpective of the study of dynamical processes on dynamical networks. The ability to tune the networks' properties is indeed crucial to understanding how each of these properties affect the dynamical processes. Generating artificial data sets that are based on empirical ones, preserve a certain number of their properties, modify oth-

ers in a tunable way, and can be extended to large sizes or long times, represents therefore a very important step in such studies.

Acknowledgments

A.B. acknowledges partial support by the FET Open project DYNANETS (number 233847) funded by the European Commission. It is a pleasure to acknowledge the SocioPatterns project (www.sociopatterns.org), partially supported by the ISI Foundation, for kindly providing the data presented in Sec. II, as well as many interesting discussions with C. Cattuto, J.-F. Pinton and W. Van den Broeck.

Appendix A: Self-consistent solution of the pairwise model

In this section we give the details of the self-consistent calculation that is able to solve for the mean-field dynamics of the pairwise interaction model. As explained in the main text, the rate equations Eqs. (5) for this model are solved together with the definition of the transition rates $\pi_{10}(t)$ and $\pi_{01}(t)$ given by Eqs. (7) by making the self-consistent assumption Eq.(8).

Inserting in the definition of $\pi_{10}(t)$ and $\pi_{01}(t)$ given by Eqs. (7) the structure of the solution of the mean-field dynamical Eq. (6) and the self-consistent assumption Eqs.(8), we get

$$\pi_{10}(t) \;=\; 2\tilde{\pi}_{01}\frac{b_1}{N}\sum_{t'=1}^{t-1}\left(\frac{t'}{N}\right)^{-\alpha_1}\left(1+\frac{t-t'}{N}\right)^{-2b_1-1} \tag{A1}$$

For large N we can evaluate (A1) by going to the continuous limit. Therefore in Eq. (A1) we substitute the sum over time steps t' with an integral over the variable $y' = t'/N$. The transition rate $\pi_{10}(y) = N\pi_{10}(t)$, that is, the average number of agents that shift from state $1 \to 0$ in the unit time $y = t/N$, can be evaluated by the following integral:

$$\begin{aligned}\pi_{10}(y) \;&=\; 2N\tilde{\pi}_{01}b_1 y^{-\alpha_1-2b_1}\int_0^1 x^{-\alpha}(1+y^{-1}-x)^{-2b_1-1}dx \\ &=\; 2N\tilde{\pi}_{01}b_1 y^{-\alpha_1} f(\alpha_1, 2b_1+1, y),\end{aligned} \tag{A2}$$

where $f(a,b,y)$ is given by

$$f(a,b,y) = y^{-(b-1)}\int_0^1 x^{-a}(1+y^{-1}-x)^{-b}dx. \tag{A3}$$

The asymptotic expansion of $f(a,b,y)$ for $y \gg 1$ is given by

$$f(a,b,y) = \frac{1}{b-1} + B(1-b,1-a)y^{1-b} + O\left(\frac{1}{y}+y^{-b}\right), \tag{A4}$$

where B is the β function. Inserting (A4) into (A2) we get

$$\pi_{10}(y) = N\tilde{\pi}_{01}y^{-\alpha_1}. \tag{A5}$$

This expression proves that the self consistent assumption given by Eq. (8) is valid. In particular since we have assumed

$$\pi_{10}(y) = N\tilde{\pi}_{10}y^{-\alpha_0}\pi_{01}(y) = N\tilde{\pi}_{01}y^{-\alpha_1}$$

these relations are consistent with the result of Eq. (A5) obtained in the limit $N \to \infty, y \gg 1$ if

$$\begin{aligned}\alpha_0 &= \alpha_1 = \alpha \\ \tilde{\pi}_{10} &= \tilde{\pi}_{01} = \tilde{\pi}.\end{aligned} \tag{A6}$$

In order to find the expression for α and $\tilde{\pi}$ we use the conservation of the total number of agents. Indeed we have

$$\sum_{t'}\Big[N_0(t,t') + N_1(t,t')\Big] = N \tag{A7}$$

Using the Eqs. (6), (A5) and (A6) and substituting in Eq. (A7) the sum over t' with an integral over the variable $x = y'/y$, we get, in the limit $N \gg 1$

$$\begin{aligned}N\tilde{\pi}y^{-\alpha}\Big[&y^{-(2b_0-1)}\int_0^1 x^{-\alpha}(1+y-x)^{-2b_0}\;dx \\ &+y^{-(2b_1-1)}\int_0^1 x^{-\alpha}(1+y-x)^{-2b_1}dx\Big] = \;N\;,\end{aligned} \tag{A8}$$

which yields

$$\tilde{\pi}y^{-\alpha}(f(\alpha,2b_0,y) + f(\alpha,2b_1,y)) = 1 \tag{A9}$$

Finally using the asymptotic expansion Eq. (A4) we get the solution given by the Eqs. (9) that we rewrite here for convenience

$$\begin{aligned}\alpha \;&=\; \max\left(0, 1-2b_1, 1-2b_0\right)\;, \\ \tilde{\pi} \;&=\; \frac{\sin\left[2\pi\min\left(b0,b1\right)\right]}{\pi}[1-\delta(\alpha,0)] \\ &\quad+\frac{(2b_0-1)(2b_1-1)}{2(b_0+b_1-1)}\delta(\alpha,0).\end{aligned} \tag{A10}$$

Appendix B: Self-consistent solution of the general model

In this appendix we solve the general model in which groups of different size are allowed and the parameter λ is arbitrary. The strategy that leads to the solution of the mean-field equation of this dynamics is essentially the same as in the pairwise model but a new phase transition occurs when $\lambda < 0.5$. The dynamical Eqs. (12) can be solved as a function of the variables $\pi_{mn}(t)$ by Eqs. (14) and Eqs. (16) assuming self-consistently that

that $\epsilon(t) = \hat{\epsilon}$ in the large time limit. In order to find the analytic solution of the mean-field dynamics it therefore important to determine the relations between the transition rates $\pi_{mn}(t)$ and the variables $N_n(t,t')$. These relations are given by

$$\begin{aligned}
\pi_{1,0}(t) &= 2\lambda \sum_{t'} \frac{N_1(t,t')}{N} p_1(t,t') \\
\pi_{n,0}(t) &= \lambda \sum_{t'} \frac{N_n(t,t')}{N} p_1(t,t'),\ n \geq 2 \\
\pi_{n+1,n}(t) &= (n+1)\lambda \sum_{t'} \frac{N_{n+1}(t,t')}{N} p_1(t,t'),\ i \geq 1 \\
\pi_{0,1}(t) &= 2 \sum_{t'} \frac{N_0(t,t')}{N} p_0(t,t') \\
\pi_{0,n}(t) &= (1-\lambda) \sum_{t'} \frac{N_{n-1}(t,t')}{N} p_1(t,t'),\ n \geq 2 \\
\pi_{n,n+1}(t) &= (n+1)(1-\lambda) \sum_{t'} \frac{N_n(t,t')}{N} p_1(t,t'),\ n \geq 1.
\end{aligned} \tag{B1}$$

The coupled Eqs. (14), (16) and (B1) can be solved by making the additional self-consistent assumptions on the transition rates $\pi_{mn}(t)$ given by

$$\pi_{mn}(y) = N\tilde{\pi}_{mn} y^{-\alpha_{mn}} \tag{B2}$$

where $y = t/N$ and $\pi_{mn}(y) = N\pi_{mn}(t)$.

Applying the same technique as in Appendix A we can prove that all the exponents $\alpha_{m,n}$ are equal and given by $\alpha_{m,n} = \alpha$. Performing straightforward calculations we get the following relations

$$\begin{aligned}
\tilde{\pi}_{n,0}(n+1) &= \lambda[\tilde{\pi}_{n-1,n} + \tilde{\pi}_{n+1,n} + \tilde{\pi}_{0,n}] \quad \text{for } n \geq 2 \\
\tilde{\pi}_{1,0} &= \lambda[\tilde{\pi}_{1,0} + \tilde{\pi}_{2,1}] \\
\tilde{\pi}_{n,0} &= (1-\lambda)\tilde{\pi}_{n-1,0} + \lambda\tilde{\pi}_{n+1,0} \quad \text{for } n \geq 3 \\
\tilde{\pi}_{2,0} &= \frac{1-\lambda}{2}\tilde{\pi}_{1,0} + \lambda\tilde{\pi}_{3,0} \\
\tilde{\pi}_{1,0} &= \lambda\tilde{\pi}_{1,0} + 2\lambda\tilde{\pi}_{2,0}.
\end{aligned} \tag{B3}$$

Therefore if the self-consistent assumption is valid, the number of agents $N_n(t,t')$ in state n since time t', is given at time t by

$$\begin{aligned}
N_0(t,t') &= \frac{\pi_{1,0}(t')}{K}\left(1 + \frac{t-t'}{N}\right)^{-b_0[2+(1-\lambda)\hat{\epsilon}]} \\
N_1(t,t') &= \frac{\pi_{1,0}(t')}{\lambda}\left(1 + \frac{t-t'}{N}\right)^{-2b_1} \\
N_n(t,t') &= \frac{(n+1)\pi_{n,0}(t')}{\lambda}\left(1 + \frac{t-t_0}{N}\right)^{-(n+1)b_1}
\end{aligned} \tag{B4}$$

where the variable K is defined by

$$K = \frac{\tilde{\pi}_{1,0}}{\sum_{n\geq 1}\tilde{\pi}_{n,0}}. \tag{B5}$$

Using the relations given by Eqs. (B3) we find

$$\tilde{\pi}_{n,0} = \frac{1}{2}\tilde{\pi}_{1,0}\left(\frac{1-\lambda}{\lambda}\right)^{n-1} \quad \text{for } n \geq 2. \tag{B6}$$

Substituting Eq. (B6) in the definition of K, Eq. (B5), we find that K is only defined for $\lambda > 0.5$. For $\lambda < 0.5$ the summation in Eq. (B5) is in fact divergent and there is a breakdown of the self-consistent assumption Eq. (B2). For $\lambda > 0.5$ we can perform the summation and we get

$$\begin{aligned}
K &= \frac{2(2\lambda-1)}{3\lambda-1}, \\
\hat{\epsilon} &= \frac{1}{2\lambda-1}.
\end{aligned} \tag{B7}$$

Finally the value of α and $\tilde{\pi}_{1,0}$ are found by enforcing the conservation law of the number of agent N

$$\sum_{t'=1}^{t}\sum_{n} N_n(t,t') = N. \tag{B8}$$

Therefore, in the large y limit $y \gg 1$ we get the solution

$$\alpha = \max\left(0, 1 - b_0\frac{3\lambda-1}{2\lambda-1}, 1-2b_1\right). \tag{B9}$$

The value of $\tilde{\pi}_{1,0}$ depends on the value assumed by α.

(1) For $\alpha = 0$, the value of $\tilde{\pi}_{1,0}$ is given by

$$\begin{aligned}
\tilde{\pi}_{1,0} = \Bigg[&\frac{1}{2(b_0 - \frac{2\lambda-1}{3\lambda-1})} \\
&+ \frac{1}{2\lambda}\sum_{n\geq 1}\frac{n+1}{(n+1)b_1 - 1}\left(\frac{1-\lambda}{\lambda}\right)^{n-1}\Bigg]^{-1}
\end{aligned} \tag{B10}$$

(2) For $\alpha = 1 - b_0\frac{3\lambda-1}{2\lambda-1}$, the value of $\tilde{\pi}_{1,0}$ is given by

$$\tilde{\pi}_{1,0} = \frac{2(2\lambda-1)}{3\lambda-1}\frac{1}{B(1-b_0\frac{3\lambda-1}{2\lambda-1}, b_0\frac{3\lambda-1}{2\lambda-1})}, \tag{B11}$$

where $B(a,b)$ indicates the Beta function.

(3) For $\alpha = 1 - 2b_1$, the value of $\tilde{\pi}_{1,0}$ is given by

$$\tilde{\pi}_{1,0} = \frac{\lambda}{B(1-2b_1, 2b_1)} \tag{B12}$$

where $B(a,b)$ indicates the Beta function.

The average coordination number is defined by

$$\langle n \rangle = \sum_{t'}^{t}\sum_{n=0}^{N} nN_n(t,t'). \tag{B13}$$

Substituting Eqs. (B4) to the definition of $\langle n \rangle$, Eq. (B13) and applying the same transformation in Eq. (A2) to evaluate the integral over t, we get

$$\langle n \rangle = \sum_{n=1}^{N}\frac{\tilde{\pi}_{n0}}{\lambda}(n+1-\delta_{n,1})y^{-\alpha}f(\alpha,(n+1)b_1,y) \tag{B14}$$

where $y = t/N$ and $f(a,b,y)$ is defined in Eq. (A3). Substituting the asymptotic expansion Eq. (A4) into (B14), we get

$$\begin{aligned}\langle n\rangle &= \sum_{n=1}^{N}\frac{\tilde{\pi}_{n0}}{\lambda}(n+1-\delta_{n,1})y^{-\alpha}\Bigg[\frac{1}{(n+1)b_1-1}\\ &+ B(1-(n+1)b_1,1-\alpha)y^{1-(n+1)b_1}\Bigg]. \qquad \text{(B15)}\end{aligned}$$

where $B(a,b)$ indicates the Beta function. In the asymtotic limit $y\to\infty$, using Eqs. (B6), (B10)-(B12) and counting only the leading terms in Eq. (B15) to compute $\langle n\rangle$ for different value of α, we can recover Eqs. (20)-(22).

Appendix C: Self-consistent solution of the heterogeneous model for $\lambda = 1$

In this appendix we show the self-consistent calculations that solve analytically the heterogeneous model with pairwise interactions.

We assume self-consistently that the transition rate $\pi_{10}^{\eta}(t)$ and $\pi_{01}^{\eta,\eta'}$ decay in time as a power-law, i.e. we assume

$$\begin{aligned}\pi_{10}^{\eta}(t) &= \Delta\eta\tilde{\pi}_{10}^{\eta}\Big(\frac{t}{N}\Big)^{-\alpha(\eta)}\\ \pi_{01}^{\eta\eta'}(t) &= \Delta\eta\Delta\eta'\tilde{\pi}_{01}^{\eta\eta'}\Big(\frac{t}{N}\Big)^{-\alpha(\eta,\eta')} \qquad \text{(C1)}\end{aligned}$$

Inserting this self-consistent assumption and the structure of the solution given by Eqs. (25) in Eqs. (26) we can evaluate $\tilde{\pi}^{\eta,\eta'}$ in the limit $N\to\infty$. Therefore we get,

$$\begin{aligned}\tilde{\pi}_{01}^{\eta\eta'}\ \ y^{\ -\alpha(\eta,\eta')} &= \frac{2N}{C(y)}\eta\tilde{\pi}_{10}^{\eta}y^{-\alpha(\eta)}f(\alpha(\eta),2\eta+1,y)\\ &\eta'\tilde{\pi}_{10}^{\eta'}y^{-\alpha(\eta')}f(\alpha(\eta'),2\eta'+1,y) \qquad \text{(C2)}\end{aligned}$$

where $f(a,b,y)$ is given by

$$f(a,b,y) = y^{-(b-1)}\int_0^1 x^{-a}(1+y^{-1}-x)^{-b}dx. \qquad \text{(C3)}$$

The asymptotic expansion to $f(a,b,y)$ for $y\gg 1$ is given by

$$f(a,b,y) = \frac{1}{b-1} + B(1-b,1-a)y^{1-b} + O\left(\frac{1}{y}+y^{-b}\right) \qquad \text{(C4)}$$

where B is the Beta function. Inserting (C4) into (C2), we get in the limit $y\gg 1$

$$\tilde{\pi}_{01}^{\eta\eta'}y^{-\alpha(\eta,\eta')} = \frac{N}{2C(y)}\tilde{\pi}_{10}^{\eta}y^{-\alpha_0(\eta)}\tilde{\pi}_{10}^{\eta'}y^{-\alpha(\eta')} \qquad \text{(C5)}$$

Similarly, inserting (C1) into the definition of $C(y)$ given by Eq. (27) we get, in the limit $y\gg 1$

$$C(y) = \frac{N}{2}\int_0^1 y^{-\alpha(\eta)}\tilde{\pi}_{10}^{\eta}d\eta \qquad \text{(C6)}$$

where we make use of the asymptotic expansion (C4). In the limit $y\gg 1$, the integral above can be calculated approximately by the saddle point method if $\tilde{\pi}_{10}^{\eta}$ changes with η much slower than $y^{-\alpha(\eta)}$. Therefore we have

$$\frac{2C(y)}{N} = \tilde{\pi}_{10}^{\eta^\star}y^{-\gamma} \qquad \text{(C7)}$$

where γ and $\eta^\star$ are given by

$$\begin{aligned}\gamma &= \min_{\eta}\alpha(\eta)\\ \eta^\star &= \text{argmin}_{\eta}\alpha(\eta). \qquad \text{(C8)}\end{aligned}$$

By comparing both sides of Eq. (C5) and using Eq. (C7) we get

$$\begin{aligned}\tilde{\pi}_{01}^{\eta\eta'} &= \frac{1}{\tilde{\pi}_{10}^{\eta^\star}}\tilde{\pi}_{10}^{\eta}\tilde{\pi}_{10}^{\eta'}\\ \alpha(\eta,\eta') &= \alpha(\eta)+\alpha(\eta')+\gamma. \qquad \text{(C9)}\end{aligned}$$

Finally, in order to fully solve the problem we impose the conservation laws of this heterogeneous model. In particular the total number of agent with value $\eta_i\in(\eta,\eta+\Delta\eta)$ is given by the following relation,

$$\sum_{t'}[N_0(t,t',\eta)+\sum_{\eta'}N_1(t,t',\eta,\eta')] = N\Delta(\eta). \qquad \text{(C10)}$$

Inserting the self-consistent anzatz Eq. (C1) for $\pi_{01}(t)$ and Eq. (C5) into Eq. (C10) we get, in the continuous limit approximation valid for $N\gg 1$,

$$\begin{aligned}\tilde{\pi}_{10}^{\eta}y^{-\alpha(\eta)} &= \Bigg[\frac{\theta(2\eta-1)}{2\eta-1}+\theta(1-2\eta)\\ &\times B(1-2\eta,1-\alpha(\eta))y^{1-2\eta}+I(\eta)\Bigg]^{-1} \qquad \text{(C11)}\end{aligned}$$

where

$$\begin{aligned}I(\eta) &= \frac{N}{2C(y)}\int_0^1\Bigg[\frac{\theta(1-\eta-\eta')}{1-\eta-\eta'}+\theta(\eta+\eta'-1)\\ &\times\ B(\eta+\eta'-1,1-\alpha(\eta'))y^{\eta+\eta'-1}\Bigg]\\ &\times\ \pi_{10}^{\eta'}y^{-\alpha(\eta')}d\eta'. \qquad \text{(C12)}\end{aligned}$$

We compute $I(\eta)$ defined in Eq. (C12) by counting the leading term only. Therefore we find

$$\alpha(\eta) = \max(0,1-2\eta,\eta-1+\gamma+D) \qquad \text{(C13)}$$

with D given by

$$D = \max_{\eta}[\eta-\alpha(\eta)]. \qquad \text{(C14)}$$

Solving the Eqs. (C13) and (C14) we get $\gamma = 0$ and $D = \frac{1}{2}$ and $\eta^\star = 1/2$. Therefore we can determine the exponent $\alpha(\eta)$ and $\alpha(\eta,\eta')$ that are given by

$$\begin{aligned}\alpha(\eta) &= \max\left(1-2\eta,\eta-\frac{1}{2}\right)\\ \alpha(\eta,\eta') &= \alpha(\eta)+\alpha(\eta'). \qquad \text{(C15)}\end{aligned}$$

Moreover the constants $\tilde{\pi}^{\eta}_{10}$ are given, in the limit $N \gg 1$ and $y \gg 1$, by

$$\tilde{\pi}^{\eta}_{10} = \begin{cases} \frac{\rho(\eta)}{B(1-2\eta,2\eta)} & \eta \leq \frac{1}{2} \\ \frac{\rho(\eta)}{B(\eta-\frac{1}{2},1)} & \eta \geq \frac{1}{2}. \end{cases} \tag{C16}$$

Solving equation (C18), let $\gamma + D \leq \frac{1}{2}$, then

$$\alpha(\eta) = \begin{cases} 1-2\eta & \eta \leq \frac{1}{2} \\ 0 & \frac{1}{2} \leq \eta \leq 1-\gamma-D \\ \eta-1+\gamma+D & \eta \geq 1-\gamma-D \end{cases} \tag{C17}$$

and

$$\eta - \alpha(\eta) = \begin{cases} 3\eta-1 & \eta \leq \frac{1}{2} \\ \eta & \frac{1}{2} \leq \eta \leq 1-\gamma-D \\ 1-\gamma-D & \eta \geq 1-\gamma-D \end{cases} \tag{C18}$$

obviously, $\gamma = 0$ and D is reached either at $\eta = \frac{1}{2}$ or $\eta = 1-\gamma-D$, so

$$D = \max(\frac{1}{2}, 1-D) \tag{C19}$$

The only solution to the above expression is $D = \frac{1}{2}$. Similarly, for $\gamma + D \geq \frac{1}{2}$,

$$\alpha(\eta) = \begin{cases} 1-2\eta & \eta \leq \frac{2-\gamma-D}{3} \\ \eta-1+\gamma+D & \eta \geq \frac{2-\gamma-D}{3} \end{cases} \tag{C20}$$

$$\eta - \alpha(\eta) = \begin{cases} 3\eta-1 & \eta \leq \frac{2-\gamma-D}{3} \\ 1-\gamma-D & \eta \geq \frac{2-\gamma-D}{3} \end{cases} \tag{C21}$$

Both γ and D are reached at $\eta = \frac{2-\gamma-D}{3}$, so

$$\begin{cases} \gamma = 1 - \frac{2(2-\gamma-D)}{3} \\ D = (2-\gamma-D) - 1 \end{cases} \tag{C22}$$

[1] S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW*. (Oxford University Press, Oxford, 2003).
[2] M. E. J. Newman, SIAM Review **45**, 167 (2003).
[3] R. Pastor-Satorras, A. Vespignani, *Evolution and structure of the Internet: A statistical physics approach*, Cambridge University Press, Cambridge (2004).
[4] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D.-U. Hwang, Physics Reports **424**, 175 (2006).
[5] G. Caldarelli, *Scale-Free Networks* (Oxford University Press, Oxford, 2007).
[6] A. Barrat, M. Barthélemy, A. Vespignani, *Dynamical processes on complex networks*, Cambridge University Press, Cambridge (2008).
[7] S. N. Dorogovtsev, A. V. Goltsev and J. F. F. Mendes, Rev. Mod. Phys. **80** 1275, (2008).
[8] This assumption goes usually under the name of annealed approximation.
[9] G. Bianconi, Phys. Lett. A **303**, 166 (2002).
[10] S. Bradde, F. Caccioli, L. Dall'Asta, G. Bianconi, Phys. Rev. Lett. **104**, 218701 (2010).
[11] J. Davidsen, H. Ebel, S. Bornholdt, Phys. Rev. Lett. **88**, 128701 (2002).
[12] M. Marsili, F Vega-Redondo, F Slanina, Proc. Natl. Acad. Sci. USA **101**, 1439 (2004).
[13] P. Holme and M. E. J. Newman, Phys. Rev. E **74**, 056108 (2006).
[14] F. Vazquez, V. M. Eguíluz, Maxi San Miguel, Phys. Rev. Lett. **100**, 108702 (2008).
[15] C. Nardini, B. Kozma, A. Barrat, Phys. Rev. Lett. **100**, 158701 (2008).
[16] B. Kozma, A. Barrat, Phys. Rev. E **77**, 016102 (2008).
[17] *Adaptive Networks: Theory, Models and Applications*, Springer/NECSI Studies on Complexity Series, Gross, T. and Sayama, H. Eds, 2008.
[18] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, C. Diot, Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking, 244 - 251 (2005).
[19] A. Scherrer, P. Borgnat, E. Fleury, J.-L. Guillaume, C. Robardet, Comp. Net. **52**, 2842 (2008).
[20] A. Gautreau, A. Barrat, M. Barthélemy, Proc. Natl. Acad. Sci. USA **106** 8847 (2009).
[21] `http://www.sociopatterns.org`
[22] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, A. Vespignani, PLoS ONE **5**(7) e11596 (2010).
[23] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, W. Van den Broeck, J. Theor. Biol. **271**, 166 (2011).
[24] M. Granovetter, AJS **78**, 1360 (1973).
[25] S. Wasserman, K. Faust, *Social Network Analysis: Methods and applications* (Cambridge University Press, Cambridge, 1994).
[26] J. M. Kumpula *et al.* Phys. Rev. Lett. **99**, 228701 (2007).
[27] N.F. Johnson *et al.*, Phys. Rev. E **79**, 066117 (2009).
[28] G. Palla, A.-L. Barabási, T. Vicsek, Nature **446**, 664 (2007).
[29] G. Chowell, J. M. Hyman, S. Eubank, C. Castillo-Chavez, Phys. Rev. E **68**, 066102 (2003).
[30] A. De Montis, M. Barthélemy, A. Chessa, A. Vespignani, Environmental Planning Journal B **34**, 905–924 (2007).
[31] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, Proc. Natl. Acad. Sci. USA **101**, 3747-3752 (2004).
[32] D. Brockmann, L. Hufnagel, T. Geisel, Nature **439**, 462-465 (2006).
[33] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J.J. Ramasco, A. Vespignani, Proc. Natl. Acad. Sci. USA **106** 21484-21489 (2009).
[34] M. C. González, C. A. Hidalgo, A.-L. Barabási, Nature **453**, 779-782 (2008).
[35] C. Song, Z. Qu, N. Blumm, A.-L. Barabási, Science **327**, 1018-1021 (2010).
[36] N. Eagle, A. Pentland, Personal and Ubiquitous Com-

puting **10**, 255-268 (2006).
[37] E. O'Neill, V. Kostakos, T. Kindberg, A. Fatah gen. Schieck, A. Penn, D. Stanton Fraser, T. Jones, Lecture Notes in Computer Science **4206**, 315 (2006).
[38] A. Pentland, *Honest Signals: how they shape our world* (MIT Press, Cambridge MA, 2008).
[39] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, J. H. Jones Proc. Natl. Acad. Sci. USA **107**, 22020-22025 (2010).
[40] H. Alani, M. Szomsor, C. Cattuto, W. Van den Broeck, G. Correndo, A. Barrat, 8th International Semantic Web Conference ISWC2009, Lecture Notes in Computer Science **5823**, 698-714(2009).
[41] J.-P. Eckmann, E. Moses, D. Sergi, Proc. Natl. Acad. Sci. USA **101** 14333 (2004).
[42] A.-L. Barabási, Nature **435**, 207 (2005).
[43] G. Kossinets, D. Watts, Science **311** 88-90 (2006).
[44] S. Golder, D. Wilkinson, B. Huberman, *Communities and technologies 2007: proceedings of the Third Communities and Technologies Conference*, Michigan State University, 2007.
[45] J. Leskovec, E. Horvitz, Proceeding of the 17th international conference on World Wide Web (2008) p 915-924, ACM New York, NY, USA.
[46] D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros, H. A. Makse, Proc. Natl. Acad. Sci. USA **106** 12640 (2009).
[47] R.D. Malmgren, D.B. Stouffer, A.S.L.O. Campanharo, L.A. Nunes Amaral, Science **325**, 1696-1700 (2009).
[48] J.F. Padgett, C.K. Ansell, Am J Sociol 98 (1993) 1259-1319.
[49] M.J. Lubbers, J.L. Molina, J. Lerner, U. Brandes, J. Avila, C. McCarty, Social Networks, 32 (2010) 91-104.
[50] P. Holme, Phys. Rev. E **71**, 046119 (2005).
[51] A. Vázquez, B. Rácz, A. Lukacs, A.-L. Barabàsi, Phys. Rev. Lett. **98**, 158702 (2007).
[52] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.-L. Barabàsi, Proc. Natl. Acad. Sci. USA **104**, 7332 (2007).
[53] R. Parshani, M. Dickison, R. Cohen, H. E. Stanley and S. Havlin, Europhys. Lett. **90**, 38004 (2010).
[54] J. Tang, S. Scellato, M. Musolesi, C. Mascolo and V. Latora, Phys. Rev. E **81**, 055101(R) (2010)
[55] S.A. Hill and D. Braha, Phys. Rev. E **82**, 046105 (2010).
[56] J. Stehlé, A. Barrat, G. Bianconi, Phys. Rev. E **81**, 035101(R) (2010).
[57] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[58] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, Y. Åberg, Nature **411**, 907 (2001).
[59] M. E. J. Newman, Proc. Natl. Acad. Sci. (USA) **98**, 404 (2001).
[60] Alain Barrat, Marc Barthélemy, and Alessandro Vespignani, Phys. Rev. Lett. **92**, 228701 (2004).
[61] G. Bianconi, Europhys. Lett. **71**, 1029 (2005).
[62] W.-X. Wang, B.-H. Wang, B. Hu, G. Yan, and Q. Ou, Phys. Rev. Lett. **94**, 188702 (2005).
[63] A. O. Hirschman, Am. Econ. Rev. **54**,761 (1964).
[64] O. C. Herfindahl, Copper costs and prices: 1870-1957, Johns Hopkins Press, Baltimore 1959.
[65] G. Bianconi, P. Pin and M. Marsili, PNAS **106**, 11433 (2009).