Preferential attachment hypergraph with high modularity

Frédéric Giroire, Nicolas Nisse, Thibaud Trolliet, Malgorzata Sulkowska

Preferential attachment hypergraph with high modularity*

Frédéric Giroire[1], Nicolas Nisse[1], Thibaud Trolliet[1], and Małgorzata Sulkowska[1,2]

[1] Université Côte d'Azur, CNRS, Inria, I3S, France
[2] Wrocław University of Science and Technology, Faculty of Fundamental Problems of Technology, Department of Fundamentals of Computer Science, Poland

Abstract. Numerous works have been proposed to generate random graphs preserving the same properties as real-life large scale networks. However, many real networks are better represented by hypergraphs. Few models for generating random hypergraphs exist and no general model allows to both preserve a power-law degree distribution and a high modularity indicating the presence of communities. We present a dynamic preferential attachment hypergraph model which features partition into communities. We prove that its degree distribution follows a power-law and we give theoretical lower bounds for its modularity. We compare its characteristics with a real-life co-authorship network and show that our model achieves good performances. We believe that our hypergraph model will be an interesting tool that may be used in many reasearch domains in order to reflect better real-life phenomena.

Keywords: Complex network, Hypergraph, Preferential attachment, Modularity.

1 Introduction

The area of complex networks concerns designing and analysing structures that model well large real-life systems. It was empirically recognised that the common ground of such structures are small diameter, high clustering coefficient, heavy tailed degree distribution and visible community structure [5]. Surprisingly, all those characteristics appear, no matter whether we investigate biological, social, or technological systems. A dynamical growth in research in this field one observes roughly since 1999 when Barabási and Albert introduced probably the most studied nowadays preferential attachment graph [2]. Their model is based on two mechanisms: growth (the graph is growing over time, gaining a new vertex and a bunch of edges at each time step) and preferential attachment (arriving vertex is more likely to attach to other vertices with high degree rather than with low degree). It captures two out of four universal properties of real networks, which are a heavy tailed degree distribution and a small world phenomenon.

* The Appendix contains proofs, comments on the model implementation and results of further experiments with a real data, that have been omitted due to lack of space.

A number of theoretical models were presented throughout last 25 years. Just to mention the mostly investigated ones: Watts and Strogatz (exhibiting small-world and high clustering properties [26]), Molloy and Reed (with a given degree sequence [20]), Chung-Lu (with a given expected degree sequence [8]), Cooper-Frieze (model of web graphs [10]), Buckley-Osthus [7] or random intersection graph (with high clustering properties and following a power-law, [4]). None of here mentioned graphs captures all the four properties listed in the previous paragraph, e.g., [26] does not have a heavy tailed degree distribution, [2] and [8] models suffer from vanishing clustering coefficient [5], almost all of them do not exhibit visible community structure, i.e., have low *modularity*.

Modularity is a parameter measuring how clearly a network may be divided into *communities*. It was introduced by Newman and Girvan in [22]. A graph has high modularity if it is possible to partition the set of its vertices into communities inside which the density of edges is remarkably higher than the density of edges between different communities. Modularity is known to have some drawbacks (for thorough discussion check [18]). Nevertheless, today it remains a popular measure and is widely used in most common algorithms for community detection [12, 3, 24]. It is well known that the real-life social or biological networks are highly modular [11, 13]. At the same time simulations show that most of existing preferential attachment models have low modularity. Good modularity properties one finds in geometric models, like spatial preferential attachment graphs [16, 15], however they use additionally a spatial metric.

Finally, almost all the up-to-date complex networks models are graph models thus are able to mirror only binary relations. In practical applications k-ary relations (co-authorship, groups of interests or protein reactions) are often modelled in graphs by cliques which may lead to a profound information loss.

Results. Within this article we propose a dynamic model with high modularity by preserving a heavy tailed degree distribution and not using a spatial metric. Moreover, our model is a random hypergraph (not a graph) thus can reflect k-ary relations. Preferential attachment hypergraph model was first introduced by Wang et al. in [25]. However, it was restricted just to a specific subfamily of uniform acyclic hypergraphs (the analogue of trees within graphs). The first rigorously studied non-uniform hypergraph preferential attachment model was proposed only in 2019 by Avin et al. [1]. Its degree distribution follows a power-law. However, our empirical results indicate that this model has a weakness of low modularity (see Section 5.2). To the best of our knowledge the model proposed within this article is the first dynamic non-uniform hypergraph model with degree sequence following a power-law and exhibiting clear community structure. We experimentally show that features of our model correspond to the ones of a real co-authorship network built upon Scopus database.

Paper organisation. Basic definitions are introduced in Sec. 2. In Sec. 3, we present a universal preferential attachment hypergraph model which unifies many existing models (from classical Barabási-Albert graph [2] to Avin et al. preferential attachment hypergraph [1]). In Sec. 4, we use it as a component in a stochastic block model to build a general hypergraph with good modularity

properties. Theoretical bounds for its modularity and experimental results on a real data are presented in Sec. 5. Further works are presented in Sec. 6.

2 Basic definitions and notation

We define a *hypergraph* H as a pair $H = (V, E)$, where V is a set of vertices and E is a set of hyperedges, i.e., non-empty, unordered multisets of V. We allow for a multiple appearance of a vertex in a hyperedge (self-loops). The degree of a vertex v in a hyperedge e, denoted by $d(v, e)$, is the number of times v appears in e. The cardinality of a hyperedge e is $|e| = \sum_{v \in e} d(v, e)$. The degree of a vertex $v \in V$ in H is understood as the number of times it appears in all hyperedges, i.e., $\deg(v) = \sum_{e \in E} d(v, e)$. If $|e| = k$ for all $e \in E$, H is said *k-uniform*.

We consider hypergraphs that grow by adding vertices and/or hyperedges at discrete time steps $t = 0, 1, 2, \ldots$. The hypergraph obtained at time t will be denoted by $H_t = (V_t, E_t)$ and the degree of $u \in V_t$ in H_t by $\deg_t(u)$. By D_t we denote the sum of degrees at time t, i.e., $D_t = \sum_{u \in V_t} \deg_t(u)$. As the hypergraph gets large, the probability of creating a self-loop can be well bounded and is quite small provided that the sizes of hyperedges are reasonably bounded.

$N_{k,t}$ stands for the number of vertices in H_t of degree k. We say that the degree distribution of a hypergraph follows a *power-law* if the fraction of vertices of degree k is proportional to $k^{-\beta}$ for some exponent $\beta \geq 1$. Formally, we will interpret it as $\lim_{t \to \infty} \mathbb{E}\left[\frac{N_{k,t}}{|V_t|}\right] \sim c \cdot k^{-\beta}$ for some positive constant c and $\beta \geq 1$. For f and g being real functions we write $f(k) \sim g(k)$ if $f(k)/g(k) \xrightarrow{k \to \infty} 1$.

Modularity measures the presence of community structure in the graph. Its definition for graphs introduced by Newman and Girvan in 2004 is given below.

Definition 1 ([22]). *Let $G = (V, E)$ be a graph with at least one edge. For a partition $\mathcal{A}$ of vertices of G define its modularity score on G as*

$$q_{\mathcal{A}}(G) = \sum_{A \in \mathcal{A}} \left(\frac{|E(A)|}{|E|} - \left(\frac{vol(A)}{2|E|} \right)^2 \right),$$

where $E(A)$ is the set of edges within A and $vol(A) = \sum_{v \in A} deg(v)$. Modularity of G is given by $q^*(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G)$.

Conventionally, a graph with no edges has modularity equal to 0. The value $\sum_{A \in \mathcal{A}} \frac{|E(A)|}{|E|}$ is called an *edge contribution* while $\sum_{A \in \mathcal{A}} \left(\frac{vol(A)}{2|E|} \right)^2$ is a *degree tax*. A single summand of the modularity score is the difference between the fraction of edges within A and the expected fraction of edges within A if we considered a random multigraph on V with the degree sequence given by G. One can observe that the value of $q^*(G)$ always falls into the interval $[0, 1)$.

Several approaches to define a modularity for hypergraphs can be found in contemporary literature. Some of them flatten a hypergraph to a graph (e.g., by replacing each hyperedge by a clique) and apply a modularity for graphs (see e.g. [21]). Others base on information entropy modularity [27]. We want to stick

to the classical definition from [22] and preserve a rich hypergraph structure, therefore we work with the definition proposed by Kamiński et al. in [17].

Definition 2 ([17]). *Let $H = (V, E)$ be a hypergraph with at least one hyperedge. For $\ell \geq 1$ let $E_\ell \subseteq E$ denote the set of hyperedges of cardinality ℓ. For a partition $\mathcal{A}$ of vertices of H define its modularity score on H as*

$$q_{\mathcal{A}}(H) = \sum_{A \in \mathcal{A}} \left(\frac{|E(A)|}{|E|} - \sum_{\ell \geq 1} \frac{|E_\ell|}{|E|} \cdot \left(\frac{vol(A)}{vol(V)} \right)^\ell \right),$$

where $E(A)$ is the set of hyperedges within A (a hyperedge is within A if all its vertices are contained in A), $vol(A) = \sum_{v \in A} deg(v)$ and $vol(V) = \sum_{v \in V} deg(v)$. Modularity of H is given by $q^*(H) = \max_{\mathcal{A}} q_{\mathcal{A}}(H)$.

A single summand of the degree tax is the expected number of hyperedges within A if we considered a random hypergraph on V with the degree sequence given by H and having the same number of hyperedges of corresponding cardinalities.

We write that an event A occurs *with high probability* (whp) if the probability $\mathbb{P}[A]$ depends on a certain number t and tends to 1 as t tends to infinity.

3 General preferential attachment hypergraph model

In this section we generalise a hypergraph model proposed by Avin et al. in [1]. Model from [1] allows for two different actions at a single time step - attaching a new vertex by a hyperedge to the existing structure or creating a new hyperedge on already existing vertices. We allow for four different events at a single time step, admit the possibility of adding more than one hyperedge at once and draw the cardinality of newly created hyperedge from more than one distribution. The events allowed at a single time step in our model H_t are: adding an isolated vertex, adding a vertex and attaching it to the existing structure by m hyperedges, adding m hyperedges, or doing nothing. The last event "doing nothing" is included since later we put H_t in a broader context of stochastic block model, where it serves as a single community. "Doing nothing" indicates a time slot in which nothing associated directly with H_t happens but some event takes place in the other part of the whole stochastic block model.

3.1 Model $\mathbf{H(H_0, p, Y, X, m, \gamma)}$

General hypergraph model H is characterized by six parameters. These are:

1. H_0 - initial hypergraph, seen at $t = 0$;
2. $\mathbf{p} = (p_v, p_{ve}, p_e)$ - vector of probabilities indicating, what are the chances that a particular type of event occurs at a single time step; we assume $p_v + p_{ve} + p_e \in (0, 1]$; additionally p_e is split into the sum of r probabilities $p_e = p_e^{(1)} + p_e^{(2)} + \ldots + p_e^{(r)}$ which allows for adding hyperedges whose cardinalities follow different distributions;

3. $Y = (Y_0, Y_1, \ldots, Y_t, \ldots)$ - independent random variables, cardinalities of hyperedges that are added together with a vertex at a single time step;
4. $X = ((X_1^{(1)}, \ldots, X_t^{(1)}, \ldots), (X_1^{(2)}, \ldots, X_t^{(2)}, \ldots), \ldots, (X_1^{(r)}, \ldots, X_t^{(r)}, \ldots))$ - r sequences of independent random variables, cardinalities of hyperedges that are added at a single time step when no new vertex is added;
5. m - number of hyperedges added at once;
6. $\gamma \geqslant 0$ - parameter appearing in the formula for the probability of choosing a particular vertex to a newly created hyperedge.

Here is how the structure of $H = H(H_0, p, Y, X, m, \gamma)$ is being built. We start with some non-empty hypergraph H_0 at $t = 0$. We assume for simplicity that H_0 consists of a hyperedge of cardinality 1 over a single vertex. Nevertheless, all the proofs may be generalised to any initial H_0 having constant number of vertices and constant number of hyperedges with constant cardinalities. 'Vertices chosen from V_t in proportion to degrees' means that vertices are chosen independently (possibly with repetitions) and the probability that any u from V_t is chosen is

$$\mathbb{P}[u \text{ is chosen}] = \frac{\deg_t(u) + \gamma}{\sum_{v \in V_t}(\deg_t(v) + \gamma)} = \frac{\deg_t(u) + \gamma}{D_t + \gamma|V_t|}.$$

For $t \geqslant 0$ we form H_{t+1} from H_t choosing only one of the following events according to $\mathbf{p}$.

- With probability p_v: Add one new isolated vertex.
- With probability p_{ve}: Add one vertex v. Draw a value y being a realization of Y_t. Then repeat m times: select $y - 1$ vertices from V_t in proportion to degrees; add a new hyperedge consisting of v and $y - 1$ selected vertices.
- With probability $p_e^{(1)}$: Draw a value x being a realization of $X_t^{(1)}$. Then repeat m times: select x vertices from V_t in proportion to degrees; add a new hyperedge consisting of x selected vertices.
- ...
- With probability $p_e^{(r)}$: Draw a value x being a realization of $X_t^{(r)}$. Then repeat m times: select x vertices from V_t in proportion to degrees; add a new hyperedge consisting of x selected vertices.
- With probability $1 - (p_v + p_{ve} + p_e)$: Do nothing.

We allow for r different distributions from which one can draw the cardinality of newly created hyperedges. Later, when H_t serves as a single community in the context of the whole stochastic block model, this trick allows for spanning a new hyperedge across several communities drawing vertices from each of them according to different distributions. This reflects some possible real-life applications. Think of an article authored by people from two different research centers. Our experimental observation is that it is very unlikely that the number of authors will be distributed uniformly among two centers. More often, one author represents one center, while the others are affiliated with the second one.

3.2 Degree distribution of $\mathbf{H(H_0, p, Y, X, m, \gamma)}$

In this section we prove that the degree distribution of $H = H(H_0, p, Y, X, m, \gamma)$ follows a power-law with $\beta > 2$. We assume that supports of random variables indicating cardinalities of hyperedges are bounded and their expectations are constant. This assumption is in accord with potential applications - think of co-authors, groups of interest, protein reactions, ect.

Theorem 1. *Consider a hypergraph $H = H(H_0, \mathbf{p}, Y, X, m, \gamma)$ for any $t > 0$. Let $i \in \{1, \ldots, r\}$. Let $\mathbb{E}[Y_t] = \mu_0$, and $\mathbb{E}[X_t^{(i)}] = \mu_i$. Moreover, let $1 \leqslant Y_t < t^{1/4}$ and $1 \leqslant X_t^{(i)} < t^{1/4}$. Then the degree distribution of H follows a power-law with*

$$\beta = 2 + \frac{\gamma \bar{V} + m \cdot p_{ve}}{\bar{D} - m \cdot p_{ve}},$$

where $\bar{V} = p_v + p_{ve}$ and $\bar{D} = m(p_{ve}\mu_0 + p_e^{(1)}\mu_1 + \ldots + p_e^{(r)}\mu_r)$ which are the expected number of vertices added per a single time step and the expected number of vertices that increase their degree in a single time step, respectively.

Sketch of proof. We prove that $\lim_{t\to\infty} \frac{\mathbb{E}[N_{k,t}]}{|V_t|} \sim \tilde{c}k^{-\beta}$ (determining the exact constant $\tilde{c}$). For this purpose we first show that it is sufficient to prove that $\lim_{t\to\infty} \frac{\mathbb{E}[N_{k,t}]}{t} \sim ck^{-\beta}$ (Lemma 4 in the Appendix). Let $\mathcal{F}_t$ be the σ-algebra associated with the probability space at time t. Let $Q_{d,k,t}$ denote the probability that a specific vertex of degree k was chosen d times to be included in new hyperedges at time t. Moreover, let Z_t be the random variable chosen at step t among $Y_t, X_t^{(1)}, \ldots, X_t^{(r)}$ according to $(p_v, p_{ve}, p_e^{(1)}, \ldots, p_e^{(r)})$. For $t \geqslant 1$ we get that $\mathbb{E}[N_{0,t}|\mathcal{F}_{t-1}] = p_v + N_{0,t-1}Q_{0,0,t}$ and when $k \geqslant 1$:

$$\mathbb{E}[N_{k,t}|\mathcal{F}_{t-1}] = \delta_{k,m}p_{ve} + \sum_{i=0}^{\min\{k, mZ_t\}} N_{k-i,t-1}Q_{i,k-i,t},$$

where $\delta_{k,m}$ is the Kronecker delta. The proof then follows from the tedious analysis of this recursive equation. □

Below we present a bunch of examples showing that our theorem generalises the results for the degree distribution of well known models.

Example 1 (Barabási-Albert graph model, [2]). In a single time step we always add one new vertex and attach it with m edges (in proportion to degrees) to existing structure. Thus $p_v = 0$, $p_{ve} = 1$, $p_e = 0$, $\bar{V} = 1$, $Y_t = 2$, $\bar{D} = 2m$, $\gamma = 0$ and we get $\beta = 2 + \frac{m}{2m-m} = 3$.

Example 2 (Chung-Lu graph model, [9]). In a single time step: we either (with probability p) add one new vertex and attach it with an edge (in proportion to degrees) to existing structure; otherwise we just add an edge (in proportion to degrees) to existing structure. Thus $p_v = 0$, $p_{ve} = p$, $p_e = 1 - p$, $\bar{V} = p$, $Y_t = 2$, $r = 1$, $X_t^{(1)} = 2$, $\bar{D} = 2$, $m = 1$, $\gamma = 0$ and we get $\beta = 2 + \frac{p}{2-p}$.

Example 3 (Avin et al. hypergraph model, [1]). In a single time step we either (with probability p) add one new vertex and attach it with a hyperedge of cardinality Y_t (in proportion to degrees) to existing structure; otherwise we just add a hyperedge of cardinality Y_t to existing structure. The assumptions on Y_t and the sum of degrees D_t are: 1. $\lim_{t\to\infty} \frac{\mathbb{E}[D_{t-1}]/t}{\mathbb{E}[Y_t]-p_{ve}} = D \in (0,\infty)$, 2. $\mathbb{E}[|\frac{1}{D_t} - \frac{1}{\mathbb{E}[D_t]}|] = o(1/t)$, 3. $\mathbb{E}\left[\frac{Y_t^2}{D_{t-1}^2}\right] = o(1/t)$. The result from [1] states that the degree distribution of the resulting hypergraph follows a power-law with $\beta = 1 + D$. Note that in our model $\lim_{t\to\infty} \frac{\mathbb{E}[D_{t-1}]/t}{\mathbb{E}[Y_t]-p_{ve}} = \frac{\bar{D}}{\bar{D}-p_{ve}}$. Setting $p_v = 0$, $p_{ve} = p$, $p_e = 1-p$, $\bar{V} = p$, $m = 1$, $\gamma = 0$ we get $\beta = 2 + \frac{p_{ve}}{\bar{D}-p_{ve}} = 1 + \frac{\bar{D}}{\bar{D}-p_{ve}} = 1 + D$.

Remark 1. Even though our result from this section may seem similar to what was obtained by Avin et al., it is easy to indicate cases that are covered by our model but not by the one from [1] and vice versa. Indeed, the model from [1] admits a wide range of distributions for Y_t. In particular, as authors underline, three mentioned assumptions hold for Y_t which is polynomial in t. This is the case not covered by our model (we upper bound Y_t by $t^{1/4}$) but we also can not think of real-life examples that would require bigger hyperedges. Whereas we can think of some natural examples that break requirements from [1] but are admissible in our model. Put $Y_t = 2$ if t is odd and $Y_t = 3$ if t is even. Then $\lim\limits_{\substack{t\to\infty \\ t \text{ - even}}} \frac{\mathbb{E}[D_{t-1}]/t}{\mathbb{E}[Y_t]-p_{ve}} = \frac{5/2}{3-p_{ve}}$ and $\lim\limits_{\substack{t\to\infty \\ t \text{ - odd}}} \frac{\mathbb{E}[D_{t-1}]/t}{\mathbb{E}[Y_t]-p_{ve}} = \frac{5/2}{2-p_{ve}}$ thus the limit $\lim_{t\to\infty} \frac{\mathbb{E}[D_{t-1}]/t}{\mathbb{E}[Y_t]-p_{ve}}$ does not exist. Whereas in our model we are allowed to put $r = 2$, $p_e^{(1)} = p_e^{(2)} = 1/2$, $X_t^{(1)} = 2$, $X_t^{(2)} = 3$ which probabilistically simulates stated example.

4 Hypergraph model with high modularity

In this section we present a new preferential attachment hypergraph model which features partition into communities. To the best of our knowledge no mathematical model so far consolidated preferential attachment, possibility of having hyperedges and clear community structure. We prove that its degree distribution follows a power-law. We denote our hypergraph by $G_t = (V_t, E_t)$. At each time step either a new vertex (*vertex-step*) or a new hyperedge (*hyperedge-step*) is added to the existing structure. The set of vertices of G_t is partitioned into r communities $V_t = C_t^{(1)} \dot\cup C_t^{(2)} \dot\cup \ldots \dot\cup C_t^{(r)}$. Whenever a new vertex is added to G_t it is assigned to the one of r communities and stays there forever.

4.1 Model $\mathbf{G(G_0, p, M, X, P, \gamma)}$

Hypergraph model G is characterized by six parameters:

1. G_0 - initial hypergraph seen at time $t = 0$ with vertices partitioned into r communities $V_0 = C_0^{(1)} \dot\cup C_0^{(2)} \dot\cup \ldots \dot\cup C_0^{(r)}$;
2. $p \in (0,1)$ - the probability of taking a vertex-step;

3. vector $M = (m_1, m_2, \ldots, m_r)$ with all m_i positive, constant and summing up to 1; m_i is the probability that a randomly chosen vertex belongs to $C_t^{(i)}$;
4. d-dimensional matrix $P_{r\times\ldots\times r}$ of hyperedge probabilities ($P_{i_1,i_2,\ldots,i_d}$ is the probability that communities $i_1, \ldots, i_d$ share a hyperedge); d is the upper bound for the number of communities shared by a single hyperedge;
5. $X = ((X_0^{(1)}, X_1^{(1)}, \ldots), (X_0^{(2)}, X_1^{(2)}, \ldots), \ldots, (X_0^{(d)}, X_1^{(d)}, \ldots))$ - d sequences of independent random variables indicating the number of vertices from a particular community involved in a newly created hyperedge;
6. $\gamma \geqslant 0$ - parameter appearing in the formula for the probability of choosing a particular vertex to a newly created hyperedge.

We build a structure of $G(G_0, p, M, X, P, \gamma)$ starting with some initial hypergraph G_0. Here G_0 consists of r disjoint hyperedges of cardinality 1. All vertices are assigned to different communities. 'Vertices are chosen from $C_t^{(i)}$ in proportion to degrees' means that vertices are chosen independently (possibly with repetitions) and the probability that any u from $C_t^{(i)}$ is chosen equals

$$\mathbb{P}[u \text{ is chosen}] = \frac{\deg_t(u) + \gamma}{\sum_{v \in C_t^{(i)}} (\deg_t(v) + \gamma)},$$

($\deg_t(v)$ is the degree of v in G_t). For $t \geqslant 0$, G_{t+1} is obtained from G_t as follows:

- With probability p add one new isolated vertex and assign it to one of r communities according to a categorical distribution given by vector M.
- Otherwise, create a hyperedge:
  - according to P select N communities (N is a random variable depending on P) that will share a hyperedge being created, say $C_t^{(i_1)}, C_t^{(i_2)}, \ldots, C_t^{(i_N)}$;
  - assign selected communities to N random variables chosen from $\{X_t^{(1)}, \ldots, X_t^{(r)}\}$ uniformly independently at random, say to $X_t^{(j_1)}, \ldots, X_t^{(j_N)}$;
  - for each $s \in \{1, \ldots, N\}$ select $X_t^{(j_s)}$ vertices from $C_t^{(i_s)}$ in proportion to degrees;
  - create a hyperedge consisting of all selected vertices.

4.2 Degree distribution of $\mathbf{G(G_0, p, M, X, P, \gamma)}$

A power-law degree distribution of G comes from the fact that each community of G behaves over time as the hypergraph model H presented in previous section. Thus the degree distribution of each community follows a power-law. For a detailed proof and experimental results on the degree distribution of a real-life co-authorship network check the Appendix.

Theorem 2. *Consider a hypergraph $G = G(G_0, p, M, X, P, \gamma)$ for all $t > 0$. Let $\mathbb{E}[X_t^{(i)}] = \mu_i$ and $1 \leqslant X_t^{(i)} < t^{1/4}$ for $i \in \{0, 1, \ldots, r\}$. Then the degree distribution of G follows a power-law with $\beta = 2 + \gamma \cdot \min_{j\in\{1,\ldots,r\}} \{\bar{V}_j/\bar{D}_j\}$, where $\bar{V}_j$ is the expected number of vertices added to $C_t^{(j)}$ at a single time step*

and $\bar{D}_j$ is the expected number of vertices from $C_t^{(j)}$ that increase their degree at a single time step. I.e.,

$$\beta = 2 + \frac{\gamma p}{(1-p)\frac{\mu_1+\ldots+\mu_r}{r}} \cdot \min_{j\in\{1,\ldots,r\}} \left\{ \frac{m_j}{s_j} \right\},$$

where s_j is the probability that by creating a new hyperedge a community j is chosen as the one sharing it.

Remark 2. The value s_j can be derived from P; it is the sum of probabilities of creating a hyperedge between $C^{(j)}$ and any other subset of communities.

5 Modularity of $\mathbf{G(G_0, p, M, X, P, \gamma)}$

In this section we give lower bounds for the modularity of $G = G(G_0, p, M, X, P, \gamma)$ in terms of the values from matrix P. We present experimental results showing the advantage in modularity of our model over the one in [1].

5.1 Theoretical results

We analyse $G(G_0, p, M, X, P, \gamma) = (V, E)$ obtained up to time t (this time we omit superscripts t). Recall that each vertex from V is assigned to one of r communities, $V = C^{(1)} \dot\cup C^{(2)} \dot\cup \ldots \dot\cup C^{(r)}$. We obtain the lower bound for modularity deriving the modularity score of the partition $\mathcal{C} = \{C^{(1)}, C^{(2)}, \ldots, C^{(r)}\}$. This choice of partition seems obvious provided that matrix P is strongly assortative, i.e., the probabilities of having an edge inside communities are all bigger than the highest probability of having an edge joining different communities. Note that what matters for the value of modularity is the total sum of degrees in each community, not the distribution of degrees. Therefore we do not use the fact that the degree distribution follows a power-law in each community and in the whole model. We just use information from matrix P. Thus, in fact, we derive the lower bound for the modularity of stochastic block model with r communities.

For $\ell \geqslant 1$ $E_\ell \subseteq E$ is the set of hyperedges of cardinality ℓ. First, we state general lower bound for the modularity of G ias a function of matrix P.

Lemma 1. *Let $G = G(G_0, p, M, X, P, \gamma)$ with the size of each hyperedge bounded by d. Let p_i be the probability that a randomly chosen hyperedge is within community $C^{(i)}$ (i.e., all vertices of a hyperedge belong to $C^{(i)}$). By s_i we denote the probability that a randomly chosen hyperedge has at least one vertex in community $C^{(i)}$. Assume also that with high probability $|E_\ell|/|E| \sim a_\ell$ for some constants $a_\ell \in [0,1]$ and $vol(V)/|E| \sim \delta$ for some constant $\delta \in (0, \infty)$. Then whp*

$$\lim_{t\to\infty} q^*(G) \geqslant \sum_{i=1}^{r} p_i - \sum_{i=1}^{r} \sum_{\ell\geqslant 1} a_\ell \left(\frac{(d-1)s_i + p_i}{\delta} \right)^\ell.$$

Remark 3. Note that for G being 2-uniform (thus simply a graph) this result simplifies significantly to $\lim_{t\to\infty} q^*(G) \geqslant \sum_{i=1}^{r} p_i - 1/4 \sum_{i=1}^{r} (s_i + p_i)^2$.

Below we state the lower bound for the modularity of G in a version in which the knowledge of the whole matrix P is not necessary. Instead we use its two characteristics: α - the probability that a randomly chosen hyperedge joins at least two different communities (may be interpreted as the amount of noise in the network) and β - the maximum value among p_i's for $i \in \{1, 2, \ldots, r\}$. The modularity of the model will be maximised for $\alpha = 0$ (when there are no hyperedges joining different communities) and $\beta = 1/r$ (when all p_i's are equal to $1/r$ thus hyperedges are distributed uniformly across communities).

Lemma 2. *By assumptions from Lemma 1 whp* $\lim_{t\to\infty} q^*(G) \geqslant 1 - \alpha - a_1 \left(\frac{d}{\delta}\right)((d-2)\alpha + 1) - \sum_{\ell \geqslant 2} a_\ell \left(\frac{d}{\delta}\right)^\ell \left((r-1)\beta^\ell + ((d-1)\alpha + \beta)^\ell\right)$, *where* $\alpha = 1 - \sum_{i=1}^{r} p_i$ *and* $\beta = \max_{i\in\{1,\ldots,r\}} p_i$.

Remark 4. For G being 2-uniform, the result simplifies to $\lim_{t\to\infty} q^*(G) \geqslant 1 - r\beta^2 - \alpha(1 + \alpha + 2\beta)$. Note that for $\alpha = 0$ and $\beta = 1/r$, this bound equals $1 - 1/r$ and is tight, i.e., it is the modularity of the graph with the same number of edges in each of its r communities and no edges between different communities.

Remark 5. Obtained bounds work well as long as the cardinalities of hyperedges do not differ too much. This is since deriving them we bound the cardinality of each hyperedge by the size of the biggest one. In particular, the bounds are very good in case of uniform hypergraphs - check experimental results below.

5.2 Experimental results

In this subsection we show how the modularity of our model G compares with Avin et al. hypergraph A [1] and with a real-life co-authorship graph R. We also check how good is our theoretical lower bound for modularity.

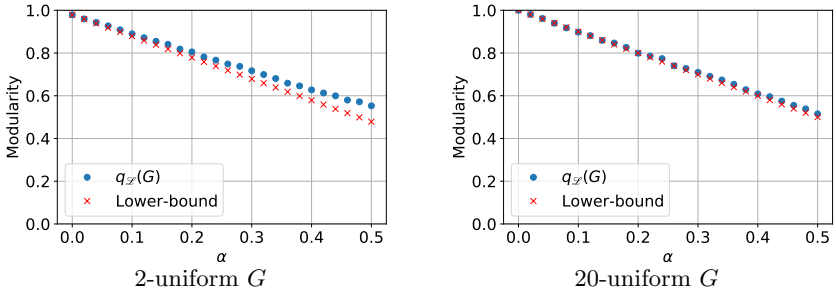


Fig. 1. Lower bound from Lemma 1 in comparison with the modularity score obtained by Leiden algorithm on simulated uniform hypergraphs G.

To get the approximation of modularity of simulated hypergraphs we used Leiden procedure [24] - a popular community detection algorithm for large networks. Calculating modularity is NP-hard [6]. Leiden is nowadays one of the best heuristics trying to find a partition maximising modularity. Therefore we treat its

outcome partition as the one whose modularity score is quite precise approximation of the modularity of graphs in question. Every presented modularity score (using Definition 2) refers to a partition returned by Leiden algorithm ran on the flattened hypergraph (i.e., a graph obtained from a hypergraph by exchanging hyperedges with cliques). We did not manage to run Leiden-like algorithm directly for hypergraphs due to their big scale and our technical limitations.

Fig. 1 shows the lower bound from Lem. 1 in comparison with the modularity of $2-$ and 20-uniform hypergraph $G(G_0, p, M, X, P, \gamma)$ on 10^4 vertices, where M is uniform and matrix P has values $(1-\alpha)/47$ (47 is the number of communities also in R) on the diagonal and the rest of probability mass spread uniformly over remaining entries. As we expected - the theoretical bound almost overlapped with the value of modularity in this case.

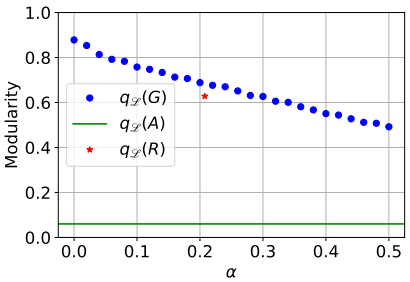


Fig. 2. Comparison of modularity between our model G, Avin et al. hypergraph A and real co-authorship hypergraph R.

To build a real-life co-authorship hypergraph R we used data downloaded from the citation database Scopus [23]. We have considered articles across all disciplines from the years 1990-2018 with at least one French co-author. Obtained hypergraph consisted of $\approx 2.2 \cdot 10^6$ nodes (authors) and $\approx 3.9 \cdot 10^6$ hyperedges (articles). Next, we implemented our model G and Avin's et al. model A using the parameters (distribution of hyperedges cardinalities, vector M, matrix P) gathered from hypergraph R. Figure 2 compares modularities of G, A, and R. For R the value α equals 0.21. Then the modularity of our model is around 0.69 which is very close to the modularity of R (≈ 0.63). The modularity of A, as A does not feature communities, is very low (≈ 0.06). Figure 2 shows also how the modularity of G changes with α and one may notice that it stays at reasonably high level even when the amount of the noise in the network grows.

6 Conclusion and Further Work

We have proved theoretically and confirmed experimentally that our model exhibits high modularity, which is rare for known preferential attachment graphs and was not present in hypergraph models so far. While our model has many parameters and may seem complicated, this general formulation allowed us to unify many results known so far. Moreover, it can be easily transformed into much simpler model (e.g., by setting some arguments trivially to 0, repeating the same distributions for hyperedges cardinalities...).

It is commonly known that many real networks present an exponential cut-off in their degree distribution. One possible reason to explain this phenomenon is that nodes eventually become inactive in the network. As further work, we will include this process in our model. The other direction of future study is making the preferential attachment depending not only on the degrees of the vertices but also on their own characteristic (generally called fitness).

References

1. Avin, C., Lotker, Z., Nahum, Y., Peleg, D.: Random preferential attachment hypergraph. In: Spezzano, F., Chen, W., Xiao, X. (eds.) ASONAM '19: International Conference on Advances in Social Networks Analysis and Mining, Vancouver, British Columbia, Canada, 27-30 August, 2019. pp. 398–405. ACM (2019)
2. Barabási, A., Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
3. Blondel, V., Guillaume, J., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech. - Theory E **2008**(10), P10008 (2008)
4. Bloznelis, M., Godehardt, E., Jaworski, J., Kurauskas, V., Rybarczyk, K.: Recent progress in complex network analysis: Models of random intersection graphs. In: Lausen, B., Krolak-Schwerdt, S., Böhmer, M. (eds.) Data Science, Learning by Latent Structures, and Knowledge Discovery. pp. 69–78. Studies in Classification, Data Analysis, and Knowledge Organization, Springer (2013)
5. Bollobás, B., Riordan, O.: Handbook of Graphs and Networks: From the Genome to the Internet. Wiley-VCH (2003), pages 1–34.
6. Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hoefer, M., Nikoloski, Z., Wagner, D.: On modularity clustering. IEEE Transactions on Knowledge and Data Engineering **20**(2), 172–188 (2008). https://doi.org/10.1109/TKDE.2007.190689
7. Buckley, P., Osthus, D.: Popularity based random graph models leading to a scale-free degree sequence. Discrete Math. **282**(1-3), 53–68 (2004)
8. Chung, F., Lu, L.: The average distances in random graphs with given expected degrees. P. Natl. Acad. Sci. USA **99**(25), 15879–15882 (2002)
9. Chung, F., Lu, L.: Complex Graphs and Networks. American Mathematical Society (2006)
10. Cooper, C., Frieze, A.: A general model of web graphs. Random Struct. Algor. **22**(3), 311–335 (2003)
11. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**, 75–174 (2010)
12. Fortunato, S., Hric, D.: Community detection in networks: A user guide. Phys. Rep. **659**, 1–44 (2016)
13. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. P. Natl. Acad. Sci. USA **99**(12), 7821–7826 (2002)
14. Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Am. Stat. Assoc. **58**(301) (1963)
15. Jacob, E., Mörters, P.: Spatial preferential attachment networks: Power laws and clustering coefficients. Ann. Appl. Probab. **25**(2), 632–662 (04 2015)
16. Kaiser, M., Hilgetagr, C.: Spatial growth of real-world networks. Phys. Rev. E **69**, 036103 (2004)
17. Kamiński, B., Poulin, V., Prałat, P., Szufel, P., Théberge, F.: Clustering via hypergraph modularity. Plos One **14**, e0224307 (Feb 2019)
18. Lancichinetti, A., Fortunato, S.: Limits of modularity maximization in community detection. Phys. Rev. E **84**, 066122 (2011)
19. Mitzenmacher, M., Upfal, E.: Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis. Cambridge University Press, USA, 2nd edn. (2017)
20. Molloy, M., Reed, B.: A critical point for random graphs with a given degree sequence. Random Struct. Algor. **6**(2/3), 161–180 (1995)
21. Neubauer, N., Obermayer, K.: Towards community detection in k-partite k-uniform hypergraphs. In: Proceedings of the NIPS 2009 Workshop on Analyzing Networks and Learning with Graphs (2009)

22. Newman, M., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E **69**, 026113 (Feb 2004)
23. Scopus: Elsevier's abstract and citation database, https://www.scopus.com/, accessed 2019-10-10
24. Traag, V., Waltman, L., van Eck, N.: From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. UK **9**(5233) (2019)
25. Wang, J., Rong, L., Deng, Q., Zhang, J.: Evolving hypernetwork model. Eur. Phys. J. B **77**, 493–498 (2010)
26. Watts, D., Strogatz, S.: Collective dynamics of small-world networks. Nature **393**, 440–442 (1998)
27. Yang, W., Wang, G., Bhuiyan, M.Z.A., Choo, K.: Hypergraph partitioning for social networks based on information entropy modularity. J. Netw. Comput. Appl. **86**, 59–71 (2017)

Appendix

Degree distribution of $\mathbf{H(H_0, p, Y, X, m, \gamma)}$

The number of vertices in H_t is a random variable following a binomial distribution. Since $|V_0| = 1$ we have $|V_t| \sim B(t, p_v + p_{ve}) + 1$. Since $|E_0| = 1$, the number of hyperedges in H_t is a random variable satisfying $|E_t| \sim mB(t, p_{ve} + p_e) + 1$.

Before we prove Theorem 1 we discuss briefly the concentration of random variables $|V_t|$ (the number of vertices at time t), D_t (the sum of degrees at time t) and $W_t = D_t + \gamma|V_t|$. We also state two technical lemmas that will be helpful later on.

Lemma 3 (Chernoff bounds, [19], Chapter 4.2). *Let $Z_1, Z_2, \ldots, Z_t$ be independent indicator random variables with $\mathbb{P}[Z_i = 1] = p_i$ and $\mathbb{P}[Z_i = 0] = 1 - p_i$. Let $\delta > 0$, $Z = \sum_{i=1}^{t} Z_i$ and $\mu = \mathbb{E}[Z] = \sum_{i=1}^{t} p_i$. Then*

$$\mathbb{P}[|Z - \mu| \geqslant \delta\mu] \leq 2e^{-\mu\delta^2/3}.$$

Corollary 1. *Since $|V_t| \sim B(t, p_v + p_{ve}) + 1$ setting $\delta = \sqrt{\frac{9 \ln t}{(p_v + p_{ve})t}}$ in Lemma 3 we get*

$$\mathbb{P}[||V_t| - \mathbb{E}[]|V_t]|| \geqslant \sqrt{9(p_v + p_{ve})t \ln t}] \leq 2/t^3.$$

Lemma 4. *If $\lim_{t\to\infty} \frac{\mathbb{E}[N_{k,t}]}{t} \sim ck^{-\beta}$ for some positive constant c then*

$$\lim_{t\to\infty} \mathbb{E}\left[\frac{N_{k,t}}{|V_t|}\right] \sim \frac{c}{p_v + p_{ve}} k^{-\beta}.$$

(*Here "$\sim$" refers to the limit by $k \to \infty$.*)

Proof. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space on which random variables $N_{k,t}$ and $|V_t|$ are defined. Thus $N_{k,t} : \Omega \to \mathbb{R}$ and $|V_t| : \Omega \to \mathbb{R}$. Let $\Omega_1 \subseteq \Omega$ denote the set of all $\omega \in \Omega$ such that $|V_t|(\omega) \in (\mathbb{E}|V_t| - \sqrt{9(p_v + p_{ve})t \ln t}, \mathbb{E}|V_t| + \sqrt{9(p_v + p_{ve})t \ln t})$. By Corollary 1 we know that $\sum_{\omega\in\Omega\setminus\Omega_1} \mathbb{P}[\omega] \leqslant 2/t^3$. Using the fact that for each ω $\frac{N_{k,t}(\omega)}{|V_t|(\omega)} \leqslant 1$ we get

$$\begin{aligned}
\mathbb{E}\left[\frac{N_{k,t}}{|V_t|}\right] &= \sum_{\omega\in\Omega} \frac{N_{k,t}(\omega)}{|V_t|(\omega)}\mathbb{P}[\omega] = \sum_{\omega\in\Omega_1} \frac{N_{k,t}(\omega)}{|V_t|(\omega)}\mathbb{P}[\omega] + \sum_{\omega\in\Omega\setminus\Omega_1} \frac{N_{k,t}(\omega)}{|V_t|(\omega)}\mathbb{P}[\omega] \\
&\leqslant \sum_{\omega\in\Omega} \frac{N_{k,t}(\omega)}{\mathbb{E}|V_t| - \sqrt{9(p_v + p_{ve})t \ln t}}\mathbb{P}[\omega] + \sum_{\omega\in\Omega\setminus\Omega_1} 1 \cdot \mathbb{P}[\omega] \\
&\leqslant \frac{\mathbb{E}[N_{k,t}]}{\mathbb{E}|V_t| - \sqrt{9(p_v + p_{ve})t \ln t}} + 2/t^3 \sim \frac{\mathbb{E}[N_{k,t}]}{(p_v + p_{ve})t}.
\end{aligned}$$

On the other hand, since $N_{k,t} \leq t$,

$$\begin{aligned}
\mathbb{E}\left[\frac{N_{k,t}}{|V_t|}\right] &\geqslant \sum_{\omega\in\Omega_1} \frac{N_{k,t}(\omega)}{|V_t|(\omega)}\mathbb{P}[\omega] \geqslant \sum_{\omega\in\Omega_1} \frac{N_{k,t}(\omega)}{\mathbb{E}|V_t| + \sqrt{9(p_v+p_{ve})t\ln t}}\mathbb{P}[\omega] \\
&= \frac{1}{\mathbb{E}|V_t| + \sqrt{9(p_v+p_{ve})t\ln t}}\left(\mathbb{E}[N_{k,t}] - \sum_{\omega\in\Omega\setminus\Omega_1} N_{k,t}(\omega)\mathbb{P}[\omega]\right) \\
&\geqslant \frac{1}{\mathbb{E}|V_t| + \sqrt{9(p_v+p_{ve})t\ln t}}\left(\mathbb{E}[N_{k,t}] - \sum_{\omega\in\Omega\setminus\Omega_1} t\cdot\mathbb{P}[\omega]\right) \\
&\geqslant \frac{\mathbb{E}[N_{k,t}]}{\mathbb{E}|V_t| + \sqrt{9(p_v+p_{ve})t\ln t}} - \frac{t\cdot 2/t^3}{\mathbb{E}|V_t| + \sqrt{9(p_v+p_{ve})t\ln t}} \\
&\sim \frac{\mathbb{E}[N_{k,t}]}{(p_v+p_{ve})t}.
\end{aligned}$$

□

Lemma 5 (Hoeffding's inequality, [14]). *Let $Z_1, Z_2, \ldots, Z_t$ be independent random variables such that $\mathbb{P}[Z_i \in [a_i, b_i]] = 1$. Let $\delta > 0$ and $Z = \sum_{i=1}^t Z_i$. Then*

$$\mathbb{P}[|Z - \mathbb{E}[Z]| \geqslant \delta] \leqslant 2\exp\left\{-\frac{2\delta^2}{\sum_{i=1}^t (a_i-b_i)^2}\right\}.$$

Lemma 6. *Let $t > 0$. Let $\mathbb{E}[Y_t] = \mu_0$, and $\mathbb{E}[X_t^{(i)}] = \mu_i$ for $i \in \{1, 2, \ldots, r\}$. Moreover, let $2 \leqslant Y_t < t^{1/4}$ and $1 \leqslant X_t^{(i)} < t^{1/4}$ for $i \in \{1, 2, \ldots, r\}$. Let $W_t = D_t + \gamma|V_t|$. Then*

$$\mathbb{P}[|W_t - \mathbb{E}[W_t]| \geq mt^{3/4}\sqrt{2\ln t}] = \mathcal{O}\left(\frac{1}{t^4}\right).$$

Proof. Our initial hypergraph consists of a single hyperedge of cardinality 1 over a single vertex thus $W_0 = \gamma + 1$. For $t \geq 1$ we can obtain W_t from W_{t-1} by adding:

1. either γ with probability p_v,
2. or $\gamma + mY_t$ with probability p_{ve},
3. or $mX_t^{(1)}$ with probability $p_e^{(1)}$,
4. or $mX_t^{(2)}$ with probability $p_e^{(2)}$,
5. or $\ldots$,
6. or $mX_t^{(r)}$ with probability $p_e^{(r)}$,
7. or 0 with probability $1 - p_v - p_{ve} - p_e$.

Thus we can express W_t as the sum of independent random variables $W_t = \gamma + 1 + Z_1 + Z_2 + \ldots + Z_t$, where $\mathbb{E}[Z_i] = \gamma\bar{V} + \bar{D}$ and $1 \leqslant Z_i \leqslant mt^{1/4} + \gamma$ for $i \in \{1, 2, \ldots, t\}$ and $\bar{D}$ and $\bar{V}$ are defined as in Theorem 1:

$$\bar{V} = p_v + p_{ve} \quad \text{and} \quad \bar{D} = m(p_{ve}\mu_0 + p_e^{(1)}\mu_1 + \ldots + p_e^{(r)}\mu_r).$$

Now, setting $\delta = mt^{3/4}\sqrt{2\ln t}$ in Hoeffding's inequality (see Lemma 5) we get

$$\mathbb{P}[|W_t - \mathbb{E}[W_t]| \geqslant mt^{3/4}\sqrt{2\ln t}] \leqslant 2\exp\left\{-\frac{4\cdot m^2\cdot t^{6/4}\cdot \ln t}{(t+1)(m\cdot t^{1/4}+\gamma)^2}\right\} = \mathcal{O}\left(\frac{1}{t^4}\right).$$

□

Lemma 7 ([9], Chapter 3.3). *Let $\{a_t\}$ be a sequence satisfying the recursive relation*

$$a_{t+1} = \left(1 - \frac{b_t}{t}\right)a_t + c_t$$

where $b_t \xrightarrow{t\to\infty} b > 0$ and $c_t \xrightarrow{t\to\infty} c$. Then the limit $\lim_{t\to\infty}\frac{a_t}{t}$ exists and

$$\lim_{t\to\infty}\frac{a_t}{t} = \frac{c}{1+b}.$$

Now we are ready to prove Theorem 1.

Proof (Theorem 1). Here we take a standard master equation approach that can be found e.g. in Chung and Lu book [9] about complex networks or Avin et al. paper [1] on preferential attachment hypergraphs.

Recall that $N_{k,t}$ denotes the number of vertices of degree k at time t. We need to show that

$$\lim_{t\to\infty}\mathbb{E}\left[\frac{N_{k,t}}{|V_t|}\right] \sim \tilde{c}k^{-\beta}$$

for some constant $\tilde{c}$ and $\beta = 2 + \frac{\gamma\bar{V}+m\cdot p_{ve}}{\bar{D}-m\cdot p_{ve}}$. However, by Lemma 4 we know that it suffices to show that

$$\lim_{t\to\infty}\frac{\mathbb{E}[N_{k,t}]}{t} \sim ck^{-\beta}$$

for some constant c.

Our initial hypergraph H_0 consists of a single hyperedge of cardinality 1 over a single vertex thus we can write $N_{0,0} = 0$ and $N_{1,0} = 1$. Now, to formulate a recurrent master equation we make the following observation for $t \geqslant 1$. The vertex v has degree k at time t if either it had degree k at time $t-1$ and was not chosen to any new hyperedge or it had degree $k-i$ at time $t-1$ and was chosen i times to new hyperedges. Note that i can be at most $\min\{k, mZ_t\}$, where Z_t represents a random variable chosen among $Y_t, X_t^{(1)}, \ldots, X_t^{(r)}$ according to $(p_v, p_{ve}, p_e^{(1)}, \ldots, p_e^{(r)})$. Additionally, at each time step a vertex of degree 0 may appear as the new one with probability p_v and a vertex of degree m may appear as the new one with probability p_{ve}. Let $\mathcal{F}_t$ be the σ-algebra associated with the probability space at time t. Let $Q_{d,k,t}$ denote the probability that a specific vertex of degree k was chosen d times to be included in new hyperedges at time t (this probability is expressed as a random variable since it depends on a specific realization of the process up to time $t-1$). Let also $W_t = D_t + \gamma|V_t|$. For $t \geqslant 1$ we get

$$\mathbb{E}[N_{0,t}|\mathcal{F}_{t-1}] = p_v + N_{0,t-1}Q_{0,0,t}$$

and when $k \geqslant 1$

$$\mathbb{E}[N_{k,t}|\mathcal{F}_{t-1}] = \delta_{k,m}p_{ve} + N_{k,t-1}Q_{0,k,t} + N_{k-1,t-1}Q_{1,k-1,t} + \sum_{i=2}^{\min\{k,mZ_t\}} N_{k-i,t-1}Q_{i,k-i,t},$$

where $\delta_{k,m}$ is the Kronecker delta. We have extracted the first two terms in the above sum since below we prove that these are the dominating terms. Taking expectation on both sides we obtain

$$\mathbb{E}[N_{0,t}] = p_v + \mathbb{E}[N_{0,t-1}Q_{0,0,t}] \tag{1}$$

and for $k \geqslant 1$

$$\begin{aligned}\mathbb{E}[N_{k,t}] = \delta_{k,m}p_{ve} + \mathbb{E}[N_{k,t-1}Q_{0,k,t}] + \mathbb{E}[N_{k-1,t-1}Q_{1,k-1,t}] \\ + \sum_{i=2}^{\min\{k,mZ_t\}} \mathbb{E}[N_{k-i,t-1}Q_{i,k-i,t}].\end{aligned} \tag{2}$$

Note that

$$\begin{aligned}Q_{0,k,t} = p_v + (1 - p_v - p_{ve} - p_e) + p_{ve}\mathbb{E}\left[\left(1 - \frac{k+\gamma}{W_{t-1}}\right)^{m(Y_t-1)} |\mathcal{F}_{t-1}\right] \\ + p_e^{(1)}\mathbb{E}\left[\left(1 - \frac{k+\gamma}{W_{t-1}}\right)^{mX_t^{(1)}} |\mathcal{F}_{t-1}\right] + \dots \\ + p_e^{(r)}\mathbb{E}\left[\left(1 - \frac{k+\gamma}{W_{t-1}}\right)^{mX_t^{(r)}} |\mathcal{F}_{t-1}\right]\end{aligned}$$

while for $i \in \{1, 2, \dots, k\}$

$$\begin{aligned}Q_{i,k-i,t} = p_{ve}\mathbb{E}\left[\binom{m(Y_t-1)}{i}\left(\frac{k-i+\gamma}{W_{t-1}}\right)^i\left(1 - \frac{k-i+\gamma}{W_{t-1}}\right)^{m(Y_t-1)-i} |\mathcal{F}_{t-1}\right] \\ + p_e^{(1)}\mathbb{E}\left[\binom{mX_t^{(1)}}{i}\left(\frac{k-i+\gamma}{W_{t-1}}\right)^i\left(1 - \frac{k-i+\gamma}{W_{t-1}}\right)^{mX_t^{(1)}-i} |\mathcal{F}_{t-1}\right] + \dots \\ + p_e^{(r)}\mathbb{E}\left[\binom{mX_t^{(r)}}{i}\left(\frac{k-i+\gamma}{W_{t-1}}\right)^i\left(1 - \frac{k-i+\gamma}{W_{t-1}}\right)^{mX_t^{(r)}-i} |\mathcal{F}_{t-1}\right].\end{aligned}$$

Now, for any random variable Z_t with constant expectation μ, independent of the σ-algebra $\mathcal{F}_{t-1}$, and such that $1 \leq Z_t < t^{1/4}$, by Bernoulli's inequality we have

$$\begin{aligned}\mathbb{E}\left[\left(1 - \frac{k+\gamma}{W_{t-1}}\right)^{mZ_t} |\mathcal{F}_{t-1}\right] &\geqslant \mathbb{E}\left[\left(1 - \frac{mZ_t(k+\gamma)}{W_{t-1}}\right) |\mathcal{F}_{t-1}\right] \\ &= 1 - \frac{m\mu(k+\gamma)}{W_{t-1}}.\end{aligned} \tag{3}$$

On the other hand (using the fact that for $x \in [0,1]$ and $n \in \mathbb{N}$ we have $(1-x)^n \leq \frac{1}{1+nx}$)

$$\begin{aligned}\mathbb{E}\left[\left(1-\frac{k+\gamma}{W_{t-1}}\right)^{mZ_t}|\mathcal{F}_{t-1}\right] &\leqslant \mathbb{E}\left[\frac{1}{1+\frac{mZ_t(k+\gamma)}{W_{t-1}}}|\mathcal{F}_{t-1}\right] \\ &= \mathbb{E}\left[1-\frac{mZ_t(k+\gamma)}{W_{t-1}+mZ(k+\gamma)}|\mathcal{F}_{t-1}\right] \\ &\leqslant \mathbb{E}\left[1-\frac{mZ_t(k+\gamma)}{W_{t-1}}+\frac{(mZ_t(k+\gamma))^2}{W_{t-1}^2}|\mathcal{F}_{t-1}\right] \\ &\leqslant 1-\frac{m\mu(k+\gamma)}{W_{t-1}}+\frac{t^{1/2}(m(k+\gamma))^2}{W_{t-1}^2},\end{aligned} \tag{4}$$

where the last inequality follows from the assumption $Z_t < t^{1/4}$. Now, let us consider the master equation (2) for $\mathbb{E}[N_{k,t}]$ term by term. We start with the expected number of vertices that had degree k at time $t-1$ and are still of degree k at time t. By (3), Lemma 6 and the fact that $N_{k,t-1} \leq t$ we get

$$\begin{aligned}\mathbb{E}[N_{k,t-1}Q_{0,k,t}] &\geqslant \mathbb{E}\left[N_{k,t-1}\left(1-\frac{(k+\gamma)m(p_{ve}(\mu_0-1)+p_e^{(1)}\mu_1+\ldots+p_e^{(r)}\mu_r)}{W_{t-1}}\right)\right] \\ &= \mathbb{E}\left[N_{k,t-1}\left(1-\frac{(k+\gamma)(\bar{D}-mp_{ve})}{W_{t-1}}\right)\right] \\ &\geqslant \mathbb{E}[N_{k,t-1}]\left(1-\frac{(k+\gamma)(\bar{D}-mp_{ve})}{\mathbb{E}[W_{t-1}]-mt^{3/4}\sqrt{2\ln t}}\right)-t\cdot\frac{1}{t^4}.\end{aligned}$$

To get the last inequality one needs to conduct calculations analogous to those from the proof of Lemma 4. By 4 and additionally using the fact that $W_{t-1} \geq 1$

$$\begin{aligned}\mathbb{E}[N_{k,t-1}Q_{0,k,t}] &\leqslant \mathbb{E}\left[N_{k,t-1}\left(1-\frac{(k+\gamma)(\bar{D}-mp_{ve})}{W_{t-1}}+\frac{t^{1/2}(p_{ve}+p_e)(m(k+\gamma))^2}{W_{t-1}^2}\right)\right] \\ &\leqslant \mathbb{E}[N_{k,t-1}]\left(1-\frac{(k+\gamma)(\bar{D}-mp_{ve})}{\mathbb{E}[W_{t-1}]+mt^{3/4}\sqrt{2\ln t}}+\frac{t^{1/2}(p_{ve}+p_e)(m(k+\gamma))^2}{(\mathbb{E}[W_{t-1}]-mt^{3/4}\sqrt{2\ln t})^2}\right) \\ &\quad+\left(t+t^{3/2}(p_{ve}+p_e)(m(k+\gamma))^2\right)\cdot\frac{1}{t^4}.\end{aligned}$$

Again, for the last inequality, proceed as in the proof of Lemma 4. Since $\mathbb{E}[W_{t-1}] = \bar{D}(t-1)+\gamma\bar{V}(t-1)$ and $\mathbb{E}[N_{k,t-1}] \leqslant t$, we obtain for fixed k

$$\mathbb{E}[N_{k,t-1}Q_{0,k,t}] = \mathbb{E}[N_{k,t-1}]\left(1-\frac{(k+\gamma)(\bar{D}-mp_{ve})}{t(\bar{D}+\gamma\bar{V})+\mathcal{O}(t^{3/4}\sqrt{\ln t})}\right)+\mathcal{O}\left(\frac{1}{\sqrt{t}}\right). \tag{5}$$

We treat $\mathbb{E}[N_{k-1,t-1}Q_{1,k-1,t}]$ similarly. On one hand we have

$$\begin{aligned}
Q_{1,k-1,t} &\geqslant p_{ve}\mathbb{E}\left[m(Y_t-1)\frac{k-1+\gamma}{W_{t-1}}\left(1-\frac{mY_t(k-1+\gamma)}{W_{t-1}}\right)|\mathcal{F}_{t-1}\right]\\
&\quad + p_e^{(1)}\mathbb{E}\left[mX_t^{(1)}\frac{k-1+\gamma}{W_{t-1}}\left(1-\frac{mX_t^{(1)}(k-1+\gamma)}{W_{t-1}}\right)|\mathcal{F}_{t-1}\right]+\dots\\
&\quad + p_e^{(r)}\mathbb{E}\left[mX_t^{(r)}\frac{k-1+\gamma}{W_{t-1}}\left(1-\frac{mX_t^{(r)}(k-1+\gamma)}{W_{t-1}}\right)|\mathcal{F}_{t-1}\right]\\
&\geqslant p_{ve}\mathbb{E}\left[m(Y_t-1)\frac{k-1+\gamma}{W_{t-1}}|\mathcal{F}_{t-1}\right]-p_{ve}\mathbb{E}\left[\frac{Y_t^2(m(k-1+\gamma))^2}{W_{t-1}^2}|\mathcal{F}_{t-1}\right]+\dots\\
&\quad + p_e^{(r)}\mathbb{E}\left[m(X_t^{(r)})\frac{k-1+\gamma}{W_{t-1}}|\mathcal{F}_{t-1}\right]-p_e^{(r)}\mathbb{E}\left[\frac{(X_t^{(r)})^2(m(k-1+\gamma))^2}{W_{t-1}^2}|\mathcal{F}_{t-1}\right]\\
&\geqslant \frac{p_{ve}m(\mu_0-1)(k-1+\gamma)}{W_{t-1}}-\frac{t^{1/2}p_{ve}(m(k-1+\gamma))^2}{W_{t-1}^2}+\dots\\
&\quad + \frac{p_e^{(r)}m\mu_r(k-1+\gamma)}{W_{t-1}}-\frac{t^{1/2}p_e^{(r)}(m(k-1+\gamma))^2}{W_{t-1}^2}\\
&= \frac{(k-1+\gamma)(\bar{D}-mp_{ve})}{W_{t-1}}-\frac{t^{1/2}(p_{ve}+p_e)(m(k-1+\gamma))^2}{W_{t-1}^2}
\end{aligned}$$

(the last inequality follows from assumptions $Y_t < t^{1/4}$ and $X_t^{(i)} < t^{1/4}$), while on the other

$$\begin{aligned}
Q_{1,k-1,t} &\leqslant p_{ve}\mathbb{E}\left[m(Y_t-1)\frac{k-1+\gamma}{W_{t-1}}|\mathcal{F}_{t-1}\right]+\dots+p_e^{(r)}\mathbb{E}\left[mX_t^{(r)}\frac{k-1+\gamma}{W_{t-1}}|\mathcal{F}_{t-1}\right]\\
&\leqslant \frac{(k-1+\gamma)(\bar{D}-mp_{ve})}{W_{t-1}}.
\end{aligned}$$

Again, by Lemma 6, the fact that $N_{k-1,t-1} \leqslant t$ and $N_{k-1,t-1}/W_{t-1} \leqslant 1$ for fixed k we get

$$\mathbb{E}[N_{k-1,t-1}Q_{1,k-1,t}] = \mathbb{E}[N_{k-1,t-1}]\left(\frac{(k-1+\gamma)(\bar{D}-mp_{ve})}{t(\bar{D}+\gamma\bar{V})+\mathcal{O}(t^{3/4}\sqrt{\ln t})}\right)+\mathcal{O}\left(\frac{1}{\sqrt{t}}\right). \tag{6}$$

The terms from equations (5) and (6) are those dominating in master equation (2). For the sum of other terms we have the following upper bound when k is fixed (the fourth inequality follows from upper bounding the sums by infinite

geometric series and the asymptotics in the last line follows from Lemma 6)

$$
\begin{aligned}
\sum_{i=2}^{\min\{k,mZ_t\}} \mathbb{E}[N_{k-i,t-1}Q_{i,k-i,t}] &\leqslant t\cdot\sum_{i=2}^{k}\mathbb{E}[Q_{i,k-i,t}] \\
&\leqslant t\cdot\mathbb{E}\left[\sum_{i=2}^{k}\left(p_{ve}\mathbb{E}\left[\binom{m(Y_t-1)}{i}\left(\frac{k-i+\gamma}{W_{t-1}}\right)^i|\mathcal{F}_{t-1}\right]\right.\right. \\
&\qquad + p_e^{(1)}\mathbb{E}\left[\binom{mX_t^{(1)}}{i}\left(\frac{k-i+\gamma}{W_{t-1}}\right)^i|\mathcal{F}_{t-1}\right]+\dots \\
&\qquad \left.\left. +p_e^{(r)}\mathbb{E}\left[\binom{mX_t^{(r)}}{i}\left(\frac{k-i+\gamma}{W_{t-1}}\right)^i|\mathcal{F}_{t-1}\right]\right)\right] \\
&\leqslant t\cdot\mathbb{E}\left[\mathbb{E}\left[\sum_{i=2}^{k}\left(p_{ve}(mY_t)^i\left(\frac{k+\gamma}{W_{t-1}}\right)^i+\dots\right.\right.\right. \\
&\qquad \left.\left.\left. +p_e^{(r)}(mX_t^{(r)})^i\left(\frac{k+\gamma}{W_{t-1}}\right)^i\right)|\mathcal{F}_{t-1}\right]\right] \\
&\leqslant t\cdot\mathbb{E}\left[\mathbb{E}\left[p_{ve}\frac{(m(k+\gamma)Y_t)^2}{W_{t-1}^2}\frac{1}{1-\frac{m(k+\gamma)Y_t}{W_{t-1}}}+\dots\right.\right. \\
&\qquad \left.\left. +p_e^{(r)}\frac{(m(k+\gamma)X_t^{(r)})^2}{W_{t-1}^2}\frac{1}{1-\frac{m(k+\gamma)X_t^{(r)}}{W_{t-1}}}|\mathcal{F}_{t-1}\right]\right] \\
&\leqslant t\cdot\mathbb{E}\left[p_{ve}\frac{(m(k+\gamma)t^{1/4})^2}{W_{t-1}^2}\frac{1}{1-\frac{m(k+\gamma)t^{1/4}}{W_{t-1}}}+\dots\right. \\
&\qquad \left. +p_e^{(r)}\frac{(m(k+\gamma)t^{1/4})^2}{W_{t-1}^2}\frac{1}{1-\frac{m(k+\gamma)t^{1/4}}{W_{t-1}}}\right] \\
&= t\cdot\mathbb{E}\left[\frac{(p_{ve}+p_e)(m(k+\gamma))^2t^{1/2}}{W_{t-1}^2}\frac{W_{t-1}}{W_{t-1}-m(k+\gamma)t^{1/4}}\right] \\
&= (p_{ve}+p_e)(m(k+\gamma))^2t^{3/2}\cdot\mathbb{E}\left[\frac{1}{W_{t-1}(W_{t-1}-m(k+\gamma)t^{1/4})}\right] \\
&\sim (p_{ve}+p_e)(m(k+\gamma))^2t^{3/2}\cdot\frac{1}{t^2}=\mathcal{O}\left(\frac{1}{\sqrt{t}}\right).
\end{aligned}
\tag{7}
$$

Plugging 5, 6 and 7 into master equation (1) and (2) we obtain

$$
\mathbb{E}[N_{0,t}] = \mathbb{E}[N_{0,t-1}]\left(1-\frac{\gamma(\bar{D}-mp_{ve})}{t(\bar{D}+\gamma\bar{V})+\mathcal{O}(t^{3/4}\sqrt{\ln t})}\right)+p_v+\mathcal{O}\left(\frac{1}{\sqrt{t}}\right) \tag{8}
$$

and

$$\mathbb{E}[N_{k,t}] = \mathbb{E}[N_{k,t-1}]\left(1 - \frac{(k+\gamma)(\bar{D}-mp_{ve})}{t(\bar{D}+\gamma\bar{V})+\mathcal{O}(t^{3/4}\sqrt{\ln t})}\right) + \mathbb{E}[N_{k-1,t-1}]\left(\frac{(k-1+\gamma)(\bar{D}-mp_{ve})}{t(\bar{D}+\gamma\bar{V})+\mathcal{O}(t^{3/4}\sqrt{\ln t})}\right) + \delta_{k,m}p_{ve} + \mathcal{O}\left(\frac{1}{\sqrt{t}}\right). \tag{9}$$

For $k \geq 0$ by L_k denote the limit

$$L_k = \lim_{t\to\infty} \frac{\mathbb{E}[N_{k,t}]}{t}.$$

First we prove that the limit L_0 exists. We apply Lemma 7 to equation (8) by setting

$$b_t = \frac{\gamma(\bar{D}-mp_{ve})}{\bar{D}+\gamma\bar{V}+\mathcal{O}(t^{3/4}\sqrt{\ln t}/t)} \quad \text{and} \quad c_t = p_v + \mathcal{O}\left(\frac{1}{\sqrt{t}}\right).$$

We get

$$\lim_{t\to\infty} b_t = \frac{\gamma(\bar{D}-mp_{ve})}{\bar{D}+\gamma\bar{V}} \qquad \text{and} \qquad \lim_{t\to\infty} c_t = p_v,$$

therefore

$$L_0 = \frac{p_v}{1+\frac{\gamma(\bar{D}-mp_{ve})}{\bar{D}+\gamma\bar{V}}} = \frac{p_v\frac{\bar{D}+\gamma\bar{V}}{\bar{D}-mp_{ve}}}{\frac{\bar{D}+\gamma\bar{V}}{\bar{D}-mp_{ve}}+\gamma}.$$

Now, we assume that the limit L_{k-1} exists and we will show by induction on k that L_k exists. Again, applying Lemma 7 to equation (9) with

$$b_t = \frac{(k+\gamma)(\bar{D}-mp_{ve})}{\bar{D}+\gamma\bar{V}+\mathcal{O}(t^{3/4}\sqrt{\ln t}/t)}$$

and

$$c_t = \frac{\mathbb{E}[N_{k-1,t-1}]}{t}\left(\frac{(k-1+\gamma)(\bar{D}-mp_{ve})}{\bar{D}+\gamma\bar{V}+\mathcal{O}(t^{3/4}\sqrt{\ln t}/t)}\right) + \delta_{k,m}p_{ve} + \mathcal{O}\left(\frac{1}{\sqrt{t}}\right)$$

we get

$$\lim_{t\to\infty} b_t = \frac{(k+\gamma)(\bar{D}-mp_{ve})}{\bar{D}+\gamma\bar{V}}$$

and

$$\lim_{t\to\infty} c_t = L_{k-1}\frac{(k-1+\gamma)(\bar{D}-mp_{ve})}{\bar{D}+\gamma\bar{V}} + \delta_{k,m}p_{ve},$$

therefore

$$L_k = \frac{L_{k-1}\frac{(k-1+\gamma)(\bar{D}-mp_{ve})}{\bar{D}+\gamma\bar{V}} + \delta_{k,m}p_{ve}}{1+\frac{(k+\gamma)(\bar{D}-mp_{ve})}{\bar{D}+\gamma\bar{V}}} = \frac{L_{k-1}(k-1+\gamma)+\delta_{k,m}p_{ve}\frac{\bar{D}+\gamma\bar{V}}{\bar{D}-mp_{ve}}}{k+\gamma+\frac{\bar{D}+\gamma\bar{V}}{\bar{D}-mp_{ve}}}. \tag{10}$$

From now on, for simplicity of notation, we put $D = \frac{\bar{D}+\gamma\bar{V}}{\bar{D}-mp_{ve}}$ thus we have

$$L_0 = \frac{p_v D}{\gamma + D} \qquad \text{and} \qquad L_k = \frac{L_{k-1}(k-1+\gamma) + \delta_{k,m} p_{ve} D}{k+\gamma+D}.$$

When $k \in \{1, 2, \ldots, m-1\}$, iterating over k gives

$$L_k = L_0 \cdot \prod_{\ell=1}^{k} \frac{\ell - 1 + \gamma}{\ell + \gamma + D} = \frac{p_v D}{\gamma + D} \prod_{\ell=1}^{k} \frac{\ell - 1 + \gamma}{\ell + \gamma + D}$$

and when $k \geqslant m$

$$\begin{aligned} L_k &= \frac{p_v D}{\gamma + D} \left(\prod_{\ell=1}^{k} \frac{\ell - 1 + \gamma}{\ell + \gamma + D} \right) + \frac{p_{ve} D}{m + \gamma + D} \left(\prod_{\ell=m+1}^{k} \frac{\ell - 1 + \gamma}{\ell + \gamma + D} \right) \\ &= \left(\frac{p_v D}{\gamma + D} \left(\prod_{\ell=1}^{m} \frac{\ell - 1 + \gamma}{\ell + \gamma + D} \right) + \frac{p_{ve} D}{m + \gamma + D} \right) \left(\prod_{\ell=m+1}^{k} \frac{\ell - 1 + \gamma}{\ell + \gamma + D} \right) \\ &= \left(\frac{p_v D}{\gamma + D} \frac{\Gamma(m+\gamma)}{\Gamma(\gamma)} \frac{\Gamma(\gamma + D + 1)}{\Gamma(m + \gamma + D + 1)} + \frac{p_{ve} D}{m + \gamma + D} \right) \\ &\quad \cdot \frac{\Gamma(m + \gamma + D + 1)}{\Gamma(m + \gamma)} \frac{\Gamma(k + \gamma)}{\Gamma(k + \gamma + D + 1)}, \end{aligned}$$

where $\Gamma(x)$ is the gamma function. Since $\lim_{k\to\infty} \frac{\Gamma(k) k^{\alpha}}{\Gamma(k+\alpha)} = 1$ for constant $\alpha \in \mathbb{R}$, we get

$$\lim_{t\to\infty} \frac{\mathbb{E}[N_{k,t}]}{t} = L_k \sim c \cdot k^{-(1+D)}$$

("$\sim$" refers to the limit by $k \to \infty$) for

$$c = p_v D \cdot \frac{\Gamma(\gamma + D)}{\Gamma(\gamma)} + p_{ve} D \cdot \frac{\Gamma(m + \gamma + D)}{\Gamma(m + \gamma)}.$$

Hence, by Lemma 4, we obtain

$$\lim_{t\to\infty} \mathbb{E}\left[\frac{N_{k,t}}{|V_t|} \right] \sim \frac{c}{p_v + p_{ve}} k^{-(1+D)}.$$

We infer that the degree distribution of our hypergraph follows power-law with

$$\beta = 1 + D = 1 + \frac{\bar{D} + \gamma\bar{V}}{\bar{D} - mp_{ve}} = 2 + \frac{\gamma\bar{V} + mp_{ve}}{\bar{D} - mp_{ve}}.$$

□

Degree distribution of $\mathbf{G(G_0, p, M, X, P, \gamma)}$

The number of vertices in G_t is a random variable satisfying $|V_t| \sim B(t, p) + r$, while for the number of hyperedges in G_t we have $|E_t| \sim B(t, 1-p) + r$. Note

that since $|V_t|$ follows a binomial distribution, Lemma 4 holds also in case of G_t if we replace $p_v + p_{ve}$ by p.

Recall that $N_{k,t}$ stands for the number of vertices in G_t of degree k. For $i \in \{1, 2, \ldots, r\}$ by $N_{k,t}^{(i)}$ we denote the number of vertices of degree k in G_t belonging to community $C_t^{(i)}$. Thus $N_{k,t} = \sum_{i=1}^{r} N_{k,t}^{(i)}$.

Lemma 8. *Consider a single community $C_t^{(j)}$ of a hypergraph G_t. Let $\mathbb{E}[X_t^{(i)}] = \mu_i$ and $1 \leqslant X_t^{(i)} < t^{1/4}$ for $i \in \{0, 1, \ldots, r\}$. Then the degree distribution of vertices from $C_t^{(j)}$ follow a power-law with*

$$\beta_j = 2 + \frac{\gamma \bar{V}_j}{\bar{D}_j}$$

where $\bar{V}_j$ is the expected number of vertices added to $C_t^{(j)}$ at a single time step and $\bar{D}_j$ is the average number of vertices from $C_t^{(j)}$ that increase their degree at a single time step, thus $\bar{V}_j = pm_j$ and $\bar{D}_j = (1-p)s_j \frac{\mu_1 + \ldots + \mu_r}{r}$, where s_j is the probability that by creating a new hyperedge a community j is chosen as the one sharing it (we obtain s_j from matrix P - see remark below).

Proof. Note that the community $C_{t+1}^{(j)}$ arises from community $C_t^{(j)}$ choosing at time t only one of the following events according to p, M and P.

- With probability pm_j: Add one new isolated vertex.
- With probability $\frac{(1-p)s_j}{r}$: Select $X_t^{(1)}$ vertices from $C_t^{(j)}$ in proportion to their degrees; these are vertices included in a newly created hyperedge, thus their degrees will increase.
- ...
- With probability $\frac{(1-p)s_j}{r}$: Select $X_t^{(r)}$ vertices from $C_t^{(j)}$ in proportion to their degrees; these are vertices included in a newly created hyperedge, thus their degrees will increase.
- With probability $1 - (pm_j + (1-p)s_j)$: Do nothing.

Now, apply Theorem 1 with $p_v = pm_j, p_{ve} = 0, p_e^{(1)} = p_e^{(2)} = \ldots = p_e^{(r)} = \frac{(1-p)s_j}{r}$ and $m = 1$. We get that the degree distribution of vertices from $C_t^{(j)}$ follow a power-law with

$$\beta_j = 2 + \frac{\gamma \bar{V}_j}{\bar{D}_j} = 2 + \frac{\gamma p m_j}{(1-p)s_j \frac{\mu_1 + \ldots + \mu_r}{r}}.$$

□

Proof (Theorem 2). We need to prove that $\lim_{t\to\infty} \mathbb{E}\left[\frac{N_{k,t}}{|V_t|}\right] \sim \tilde{c}k^{-\beta}$ for some constant $\tilde{c}$ and β as in the statement of theorem. By Lemma 4 we know that it suffices to show $\lim_{t\to\infty} \frac{\mathbb{E}[N_{k,t}]}{t} \sim ck^{-\beta}$ for some constant c. By Lemma 4 and Lemma 8 we write

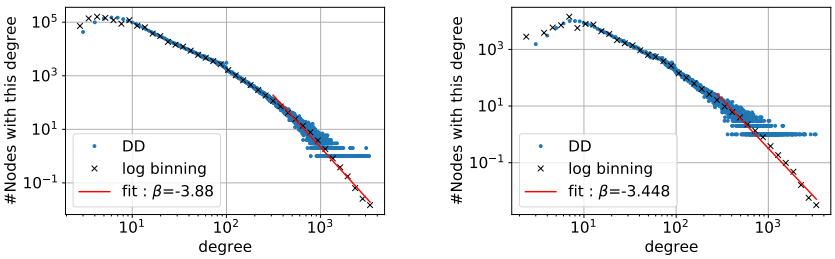
$$\begin{aligned}\lim_{t\to\infty} \frac{\mathbb{E}[N_{k,t}]}{t} &= \lim_{t\to\infty} \frac{\mathbb{E}[N_{k,t}^{(1)}]}{t} + \lim_{t\to\infty} \frac{\mathbb{E}[N_{k,t}^{(2)}]}{t} + \ldots + \lim_{t\to\infty} \frac{\mathbb{E}[N_{k,t}^{(r)}]}{t} \\ &\sim c_1 k^{-\beta_1} + c_2 k^{-\beta_2} + \ldots + c_r k^{-\beta_r}\end{aligned}$$

for some constants $c_1, \ldots, c_r$ and $\beta_j = 2 + \frac{\gamma \bar{V}_j}{\bar{D}_j}$. Thus $\lim_{t\to\infty} \frac{\mathbb{E}[N_{k,t}]}{t} \sim ck^{-\beta}$, where

$$\beta = \min_{j\in\{1,\ldots,r\}} \{\beta_j\} = 2+\gamma \cdot \min_{j\in\{1,\ldots,r\}} \left\{ \frac{\bar{V}_j}{\bar{D}_j} \right\} = 2+\frac{\gamma p}{(1-p)\frac{\mu_1+\ldots+\mu_r}{r}} \cdot \min_{j\in\{1,\ldots,r\}} \left\{ \frac{m_j}{s_j} \right\}.$$

□

In Figure 3 we present log-log plots of a power-law distribution fitted to the degree distribution (DD) of a real-life co-authorship hypergraph R. R is the same as in Section 5. Thus it is built upon Scopus database, consists of $\approx 2.2 \cdot 10^6$ nodes (authors) and $\approx 3.9 \cdot 10^6$ hyperedges (articles). Left chart presents the degree distribution of the whole R while the right one refers only to the biggest community of R found by Leiden algorithm. One can observe the power-law behaviour in both cases.



Graph R The biggest community of R

Fig. 3. A power-law distribution fitted to the degree distributions.

Let us also make one remark about the implementation of matrix P.

Remark 6. Observe that storing hyperedge probabilities in d-dimensional matrix P we use much more space than we actually should. The same probabilities may repeat many times in P. E.g., when $d = 2$ we get 2-dimensional symmetric matrix P such that $\sum_{i=1}^{r} \sum_{j=1}^{i} p_{ij} = 1$ and the the probability of creating hyperedge between two distinct communities $C^{(i)}$ and $C^{(j)}$ is in matrix P doubled - as p_{ij} and p_{ji}. If we allow for bigger hyperedges it may be repeated much more times. In fact we need to store at most $2^r - 1$ different probabilities (one for each nonempty subset of the set of communities) while in P we store d^r values (in particular, if $d = r$ we store r^r instead $2^r - 1$ values). Nevertheless, for formal proofs this notation is convenient thus we use it at the same time underlining that implementation may be done much more space efficiently.

Modularity

Proof (Lemma 1). Let $\mathcal{C} = \{C^{(1)}, C^{(2)}, \ldots, C^{(r)}\}$. Let also q denote the probability of adding a new hyperedge in a single time step (hence $q = 1 - p$, referring

to notation from Section 4). Thus with high probability $|E| \sim t \cdot q$ (where '$\sim$' refers to the limit by $t \to \infty$). By Definition 2 we write

$$q^*(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G) \geqslant q_{\mathcal{C}}(G) = \sum_{i=1}^{r} \left(\frac{|E(C^{(i)})|}{|E|} - \sum_{\ell \geqslant 1} \frac{|E_\ell|}{|E|} \left(\frac{vol(C^{(i)})}{vol(V)} \right)^\ell \right).$$

We obtain that with high probability

$$q_{\mathcal{C}}(G) \sim \sum_{i=1}^{r} \left(\frac{t \cdot q \cdot p_i}{t \cdot q} - \sum_{\ell \geqslant 1} a_\ell \left(\frac{vol(C^{(i)})}{t \cdot q \cdot \delta} \right)^\ell \right).$$

Note that if at a certain time step appears a hyperedge with all vertices contained in $C^{(i)}$, which happens with probability $q \cdot p_i$, it adds up at most d to $vol(C^{(i)})$. If at a certain time step appears a hyperedge joining at least 2 communities with at least one vertex in $C^{(i)}$, which happens with probability $q(s_i - p_i)$, it adds up at most $d-1$ to $vol(C^{(i)})$. Thus we get that with high probability

$$\begin{aligned} \lim_{t\to\infty} q^*(G) &\geqslant \sum_{i=1}^{r} p_i - \sum_{i=1}^{r} \sum_{\ell \geqslant 1} a_\ell \left(\frac{t \cdot q \cdot (dp_i + (d-1)(s_i - p_i))}{t \cdot q \cdot \delta} \right)^\ell \\ &= \sum_{i=1}^{r} p_i - \sum_{i=1}^{r} \sum_{\ell \geqslant 1} a_\ell \left(\frac{(d-1)s_i + p_i}{\delta} \right)^\ell. \end{aligned}$$

□

Proof (Lemma 2). Let $\mathcal{C} = \{C^{(1)}, C^{(2)}, \ldots, C^{(r)}\}$ and for $i \in \{1, 2, \ldots, r\}$ let $\tilde{s}_i$ be the probability that a randomly chosen hyperedge joins at least two communities and $C^{(i)}$ is one of them. Note that for s_i defined as in Lemma 1 (i.e., the probability that a randomly chosen hyperedge has at least one vertex in $C^{(i)}$) we get $s_i = \tilde{s}_i + p_i$. By Lemma 1 we get that with high probability

$$\begin{aligned} \lim_{t\to\infty} q^*(G) &\geqslant \sum_{i=1}^{r} p_i - \sum_{i=1}^{r} \sum_{\ell \geqslant 1} a_\ell \left(\frac{(d-1)\tilde{s}_i + dp_i}{\delta} \right)^\ell \\ &= (1-\alpha) - \sum_{\ell \geqslant 1} \frac{a_\ell}{\delta^\ell} \sum_{i=1}^{r} ((d-1)\tilde{s}_i + dp_i)^\ell \\ &= (1-\alpha) - \frac{a_1}{\delta} \left((d-1) \sum_{i=1}^{r} \tilde{s}_i + d \sum_{i=1}^{r} p_i \right) - \sum_{\ell \geqslant 2} \frac{a_\ell}{\delta^\ell} \sum_{i=1}^{r} ((d-1)\tilde{s}_i + dp_i)^\ell. \end{aligned} \tag{11}$$

Now, by r_k denote the probability that a randomly chosen hyperedge joins exactly k communities. Note that

$$\sum_{i=1}^{r} \tilde{s}_i = 2r_2 + 3r_3 + \ldots + dr_d \leqslant d(1 - \sum_{i=1}^{r} p_i) = d\alpha. \tag{12}$$

Thus

$$\begin{aligned}\frac{a_1}{\delta}\left((d-1)\sum_{i=1}^{r}\tilde{s}_i + d\sum_{i=1}^{r}p_i\right) &\leqslant \frac{a_1}{\delta}\left((d-1)d\alpha + d(1-\alpha)\right)\\ &= a_1\left(\frac{d}{\delta}\right)((d-2)\alpha+1).\end{aligned} \tag{13}$$

Moreover,

$$\begin{aligned}\sum_{\ell\geqslant 2}\frac{a_\ell}{\delta^\ell}\sum_{i=1}^{r}((d-1)\tilde{s}_i + dp_i)^\ell &= \sum_{\ell\geqslant 2}\frac{a_\ell}{\delta^\ell}\sum_{i=1}^{r}\sum_{k=0}^{\ell}\binom{\ell}{k}((d-1)\tilde{s}_i)^k(dp_i)^{l-k}\\ &= \sum_{\ell\geqslant 2}\frac{a_\ell}{\delta^\ell}\sum_{k=0}^{\ell}\binom{\ell}{k}(d-1)^k d^{l-k}\sum_{i=1}^{r}\tilde{s}_i^k p_i^{l-k}\\ &\leqslant \sum_{\ell\geqslant 2}\frac{a_\ell}{\delta^\ell}\sum_{k=0}^{\ell}\binom{\ell}{k}(d-1)^k(d\beta)^{l-k}\sum_{i=1}^{r}\tilde{s}_i^k\\ &= \sum_{\ell\geqslant 2}\frac{a_\ell}{\delta^\ell}\left(r(d\beta)^\ell + \sum_{k=1}^{\ell}\binom{\ell}{k}(d-1)^k(d\beta)^{l-k}\sum_{i=1}^{r}\tilde{s}_i^k\right)\\ &\leqslant \sum_{\ell\geqslant 2}\frac{a_\ell}{\delta^\ell}\left(r(d\beta)^\ell + \sum_{k=1}^{\ell}\binom{\ell}{k}(d-1)^k(d\beta)^{l-k}\left(\sum_{i=1}^{r}\tilde{s}_i\right)^k\right).\end{aligned}$$

Next, by (12) we get

$$\begin{aligned}\sum_{\ell\geqslant 2}\frac{a_\ell}{\delta^\ell}\sum_{i=1}^{r}((d-1)\tilde{s}_i + dp_i)^\ell &\leqslant \sum_{\ell\geqslant 2}\frac{a_\ell}{\delta^\ell}\left(r(d\beta)^\ell + \sum_{k=1}^{\ell}\binom{\ell}{k}(d-1)^k(d\beta)^{l-k}(d\alpha)^k\right)\\ &= \sum_{\ell\geqslant 2}\frac{a_\ell}{\delta^\ell}\left((r-1)(d\beta)^\ell + \sum_{k=0}^{\ell}\binom{\ell}{k}((d-1)d\alpha)^k(d\beta)^{l-k}\right)\\ &= \sum_{\ell\geqslant 2}\frac{a_\ell}{\delta^\ell}\left((r-1)(d\beta)^\ell + ((d-1)d\alpha + d\beta)^\ell\right)\\ &= \sum_{\ell\geqslant 2}a_\ell\left(\frac{d}{\delta}\right)^\ell\left((r-1)\beta^\ell + ((d-1)\alpha+\beta)^\ell\right).\end{aligned} \tag{14}$$

Finally, plugging (13) and (14) to (11) we get that with high probability

$$\begin{aligned}&\lim_{t\to\infty} q^*(G) \geqslant\\ &\quad\geqslant 1-\alpha-a_1\left(\frac{d}{\delta}\right)((d-2)\alpha+1) - \sum_{\ell\geqslant 2}a_\ell\left(\frac{d}{\delta}\right)^\ell\left((r-1)\beta^\ell + ((d-1)\alpha+\beta)^\ell\right).\end{aligned}$$

□