Random Preferential Attachment Hypergraphs

Chen Avin, Zvi Lotker, Yinon Nahum and David Peleg

Abstract In the future, analysis of social networks will conceivably move from graphs to hypergraphs. However, theory has not yet caught up with this type of data organizational structure. By introducing and analyzing a general model of *preferential attachment hypergraphs*, this paper makes a step towards narrowing this gap. We consider a random preferential attachment model $H(p,Y)$ for network evolution that allows arrivals of both nodes and hyperedges of random size. At each time step t, two possible events may occur: (1) [vertex arrival event:] with probability $p > 0$ a new vertex arrives and a new hyperedge of size Y_t, containing the new vertex and $Y_t - 1$ existing vertices, is added to the hypergraph; or (2) [hyperedge arrival event:] with probability $1 - p$, a new hyperedge of size Y_t, containing Y_t existing vertices, is added to the hypergraph. In both cases, the involved existing vertices are chosen at random according to the preferential attachment rule, i.e., with probability proportional to their degree, where the degree of a vertex is the number of edges containing it. Denoting the total degree in the hyper graph by $D_t = D_0 + \sum_{i=0}^{t-1} Y_i$, we allow $Y_t \geq 1$ to be any integer-valued random variable satisfying (i) $\lim_{t\to\infty} \frac{\mathbb{E}[D_t]/t}{\mathbb{E}[Y_t]-p} = \Gamma \in (0,\infty)$, (ii) $\mathbb{E}\left[|1/D_t - 1/\mathbb{E}[D_t]|\right] = o(1/t)$ and (iii) $\mathbb{E}\left[Y_t^2/D_t^2\right] = o(1/t)$. Furthermore, if Y_t is either deterministic (i.e., not random) or satisfies $\mathbb{E}\left[Y_t^2\right] = o(t)$, assumptions (ii) and (iii) can be omitted. We prove that the $H(p,Y)$ model generates *power law networks*, i.e., the expected fraction of nodes with degree k is proportional to $k^{-\beta}$, where $\beta = 1 + \Gamma$. This extends the special case of preferential attachment graphs, where $Y_t = 2$ for every t, yielding $\beta = 1 + \frac{2}{2-p}$. Therefore, our results show that the exponent of the degree distribution is sensitive to whether one considers the structure of a social network to be a hypergraph or a graph. We discuss, and provide examples for, the implications of these considerations.

Chen Avin and Zvi Lotker
Ben Gurion University of the Negev, Be'er-Sheva, Israel e-mail: {avin,zvilo}@cse.bgu.ac.il

Yinon Nahum and David Peleg
Weizmann Institute of Science, Rehovot, Israel e-mail: {yinon.nahum,david.peleg}@weizmann.ac.il

1 Introduction

Random structures have proved to be an extremely useful concept in many disciplines, including mathematics, physics, economics, and communication systems. Examining the typical behavior of random instances of a structure allows us to understand fundamental properties of the structure itself.

In the context of *graph* structures, the foundations of random graph theory were laid in a seminal paper by Erdős and Rényi in the late 1950's [8]. Subsequently, several alternative models for random structures, suitable for different kinds of applications, have been suggested. One of the most important alternative models is the *preferential attachment* (PA) model [14, 2], which was found to be particularly suitable for describing a variety of phenomena in nature, such as the "rich get richer" phenomenon, which cannot be adequately simulated within the original Erdős-Rényi model. It has been shown that the preferential attachment model captures some universal properties of real-world social networks and complex systems, such as heavy tail degree distribution and the "small world" phenomenon [12].

While graphs are extremely versatile and useful structures for representing interrelations among entities, one of their limitations is that they only capture dyadic (or binary) relations. In real-life, however, many natural, physical, and social phenomena involve k-ary relations for $k > 2$, or even relations of variable arity. For example, collaborations among researchers, as manifested through joint coauthorships of scientific papers, may be better represented using hyperedges instead of edges. Figure 1(a) depicts the hypergraph representation for coauthorship relations on four papers: paper 1 authored by $\{a,b,e,f\}$, paper 2 by $\{a,c,d,g\}$, paper 3 by $\{b,c,d\}$ and paper 4 by $\{e,f\}$. Likewise, wireless communication networks [1] or social relations captured by photos which appear on social media also give rise to hyperedges in a natural way [17]. Affiliation models [11, 13], which are a popular model for social networks, are commonly interpreted as bipartite graphs, when, in fact, they may sometimes be represented more conveniently as hypergraphs. For example, Figure 1(b) presents the bipartite graph representation of the hypergraph H of Figure 1(a). While there have been attempts to reduce hypergraphs to graphs by converting every hyperedge to a clique, as this paper will show, this indirect analysis of hypergraphs can lead to inaccuracies. Sometimes, it is only possible (or conve-

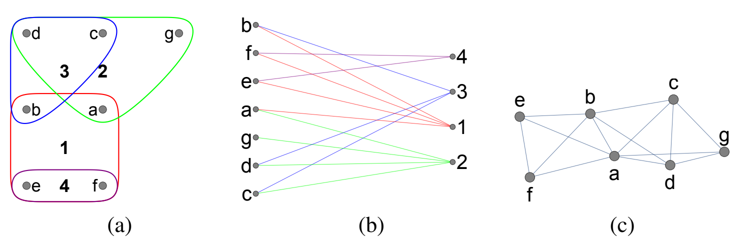

Fig. 1 (a) A hypergraph H with 7 nodes and 4 edges. (b) A bipartite graph representation of H. (c) The observed graph $G(H)$: every hyperedge is replaced with a clique.

nient) to access the *observed graph* $G(H)$ of the original hypergraph H, namely, only the pairwise relation between players is visible or given (see Figure 1(c)). In some cases, this structure may be sufficient for the application at hand. However, in many cases (e.g., when studying degree distribution), direct analysis of the original hypergraph is needed for more accurate results. In order to deal with the latter cases, we tackle the more precise, albeit somewhat harder, analysis of hypergraphs.

The study of hypergraphs, and in particular random hypergraph models, has its roots in a 1976 paper by Erdős and Bollobas [3], which offers a model analogous to the Erdős-Rényi random graph model [8]. Several interesting properties of the evolution of random hypergraphs in this model were recently studied [6, 7, 10].

A similar transition (from graphs to hypergraphs) for the preferential attachment model was first studied by Wang et al. in [15], which defined a basic evolving hypergraph model with vertex arrival events and constant-size hyperedges. Specifically, at every step, $m-1$ new nodes enter the network and a new edge of size m is added containing the $m-1$ new nodes and one existing node chosen by preferential attachment, i.e., with probability proportional to its degree. Analyzing the degree distribution, it was shown that the resulting hypergraph follows a power law with exponent $m+2$. Note, however, that the model of [15] generates restricted hypergraphs. First, they are *acyclic*, since hyperedges cannot be added between existing nodes. Second, the hyperedges are uniform, i.e., of the same size. Third, every two nodes share at most one hyperedge. Hence the model results in a limited class of uniform hypergraphs that is analogous to the class of *trees* within the family of graphs.

Our work was motivated by the fact that most actual aplications where hypergraph structures come into play do not fall in the limited class of [15]. Aiming to model a larger set of networks, we extend the $G(p)$ evolving graph model of Chung and Lu [4], which allows edge arrivals, to include hyperedges of random hyperedge size. Thus, we allow more complex structures for the resulting hypergraphs, including cycles and nonuniformity.

The main technical contribution of this paper is an analysis of the degree distribution of random preferential attachment hypergraphs, showing that they possess heavy tail degree distribution properties, similar to those of random preferential attachment graphs. However, our results also show that the exponent of the degree distribution of a random preferential attachment model is sensitive to whether one considers the structure to be a hypergraph or a graph. In fact, in the setting of hyperedges which grow in size, namely $Y_t \to \infty$, the exponent can drop below 2.

The model proposed here extends Chung and Lu's model [4] by supporting hypergraphs, and moreover, allowing hyperedges of random size. The process starts with an initial hypergraph, and at each time step t, a new hyperedge of random size joins the network. With probability p, this new hyperedge includes a new node and possibly other existing nodes, i.e., nodes which have arrived before time t, and with probability $1-p$, all the nodes of the new arriving hyperedge are existing nodes. Our model allows the hyperedge sizes to be random (with some restrictions), and the existing nodes of each hyperedge are selected randomly according to the preferential attachment rule, namely, with probability proportional to their degree.

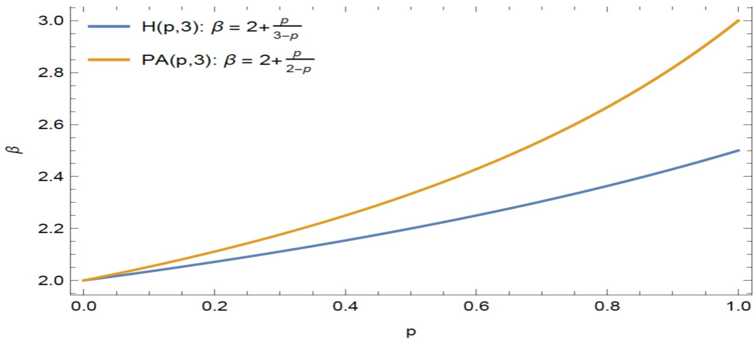

Fig. 2 The exponent β of a preferential attachment (PA) graph and a 3-uniform hypergraph as a function of p (the probability of a new vertex arrival event). In graphs, β is between 2 and 3, whereas in 3-hypergraphs β is between 2 and 2.5.

We show that the degree distribution of the resulting hypergraph follows a power law, i.e., the fraction of nodes with degree k is proportional to $k^{-\beta}$, with exponent

$$\beta = 1 + \Gamma\ , \quad \text{where} \quad \Gamma = \lim_{t\to\infty} \frac{\mathbb{E}[D_t]/t}{\mathbb{E}[Y_t] - p} \in (0,\infty)\ .$$

This naturally extends the known result for the Chung and Lu's model for graphs. Graphs are a special case of hypergraphs where hyperedges are always of size 2. In [4], Chung and Lu showed that for graphs in their model $\beta = 1 + \frac{2}{2-p}$, even when more than one edge (of size 2) is added to the network as each time step. Fig 2 illustrates the difference in β between a hypergraph model with hyperedges of size 3 and a graph model with 3 new edges at each time step as a function of p, the vertex arrival event probability.

Hence our results indicate that the seemingly natural approach of studying a hypergraph by looking at its observed graph might entail some inherent inaccuracies, and one should exercise care when applying it. In particular, it matters if the observed graph was generated by a graph or by a hypergraph evolution mechanism, since the two models generate observed graphs with *different* degree distributions.

This paper makes a step towards narrowing the gap between theory and practice in complex systems. Given a hypergraph system, it has long been possible to study it *empirically* in order to calculate its degree distribution, check whether or not it is heavy tailed, compute the exponent of the power law, and so on. However, the available *theoretical* models fell short of enabling us to accomplish these tasks analytically. The approach presented in this paper provides a useful technique toward a better theoretical understanding of these complex systems.

Related work. As a reference point, we consider the random preferential attachment graph model of Chung and Lu [4]. In that model, starting from an initial graph G_0, at any time step one of the following two possible events occur: (1) a *vertex arrival event*, occurring with probability p, where a new *vertex* joins the network and selects its neighbor from among the existing nodes via preferential attachment, or (2) an *edge arrival event*, occurring with probability $1-p$, where a new *edge* joins the network and selects its two endpoints from among the existing nodes via preferential attachment.

It is shown in [4] that the degree distribution of the random preferential attachment graph follows a power law, i.e., the probability of a random vertex to be of degree k is proportional to $k^{-\beta}$, with exponent $\beta = 1 + \frac{2}{2-p}$. A similar result can be shown in a setting where, at each time step, d edges join the graph instead of only one hyperedge (in either a vertex arrival event or an edge arrival event) [12]. This result holds even if at each step a random number of edges join the network, as long as the number of new edges has constant expectation and bounded variance.

The rest of the paper is organized as follows. Sect. 2 describes the model in detail and states our main results. Sect. 3 offers a discussion and both simulated and real-data examples. Sect. 5 presents conclusions and future questions. Sect. 4 gives a proof of the main theorem.

2 Model and results

Preliminaries. Given a set V and an integer $k \geq 1$, let $V^{(k)}$ be the set of all unordered vectors (or multisets) of k elements from V. A finite undirected *graph* G is an ordered pair (V,E) where V is a set of n *vertices* and $E \subseteq V^{(2)}$ is the set of graph *edges* (unordered pairs from V, including self-loops).

A hypergraph $\mathscr{H}$ is an ordered pair $(V,\mathscr{E})$, where V is a set of n vertices and $\mathscr{E} \subseteq \bigcup_{i=2}^{n} V^{(i)}$ is a set of *hyperedges* connecting the vertices (including self-loops, i.e., multiple appearance of a vertex in a hyperedge). When all hyperedges have the same cardinality k, the hypergraph is said to be *k-uniform*. A graph is thus simply a 2-uniform hypergraph. The *degree* of a vertex v in edge e, $d(v,e)$, is the number of times that v appears in e. The degree $d(v)$ of a vertex v is the number of time it appears in all hyperedges (counting multiplicities), i.e., $d(v) = \sum_{e\in\mathscr{E}} d(v,e)$. $\mathscr{H}$ is d-*regular* if every vertex has degree d.

In the classical preferential attachment graph model [2], the evolution process starts with an arbitrary finite initial network G_0, which is usually set to a single vertex with a self-loop. Then this initial network evolves in time, with G_t denoting the network just before time step t. Let $d_t(x)$ denote the degree of vertex x in G_t. In every time step t, a new vertex v enters the network. On arrival, v attaches itself to an existing vertex u chosen at random with probability proportional to $d_t(u)$, i.e., the *preferential attachment* rule is: $\mathbb{P}[u \text{ is chosen}] = \frac{d_t(u)}{\sum_{w\in G_t} d_t(w)}$.

The Preferential Attachment Hypergraph Model. Similar to the classical preferential attachment graph model [4], the evolution of a hypergraph in our model occurs along a discrete time axis, with one event occurring at each time step. We consider two types of possible events on the hypergraph at time t: (1) a *vertex ar-*

rival event, which involves adding a new vertex along with a new hyperedge, and (2) a *hyperedge arrival* event, where a new hyperedge is added.

We consider a nonuniform, random hypergraph where we allow multiple appearance of a vertex in a hyperedge. Similar to [4], our preferential attachment model, $H(H_0, p, Y)$, has three parameters:

1. A probability $0 < p \leq 1$ for vertex arrival events.
2. An initial hypergraph H_0 given at time 0.
3. A sequence of random independent (possibly differently distributed) integer variables $Y = (Y_0, Y_1, Y_2, \dots)$, for $Y_t \geq 1$, which determine the cardinality of the new hyperedge arriving at time t.

We next describe the process by which the random hypergraph $H(H_0, p, Y)$ grows in time. Start with the initial hypergraph H_0 at time 0. At time $t \geq 0$, the graph H_{t+1} is formed from H_t as follows.

- Draw a random size Y_t for the new hyperedge, independently of H_t.
- With probability p, add a new vertex u to V, select $Y_t - 1$ vertices from H_t (possibly with repetitions) independently according to the preferential attachment rule in H_t, and form a new hyperedge e that includes u and the $Y_t - 1$ selected vertices.
- Else, with probability $1 - p$, we select Y_t vertices from H_t (possibly with repetitions) independently according to the preferential attachment rule in H_t, and form a new hyperedge e that includes the Y_t selected vertices.

We make some assumptions on Y_t, the size of the hyperedge added at time t. Namely, denoting the total degree at time t by $D_t = \sum_{v \in H_t} d_t(v) = D_0 + \sum_{i=0}^{t-1} Y_i$, we make the following three assumptions:

$$\lim_{t \to \infty} \frac{\mathbb{E}[D_t]/t}{\mathbb{E}[Y_t] - p} = \Gamma \in (0, \infty) , \tag{1}$$

$$\mathbb{E}\left[\left|\frac{1}{D_t} - \frac{1}{\mathbb{E}[D_t]}\right|\right] = o(1/t) , \tag{2}$$

$$\mathbb{E}\left[\frac{Y_t^2}{D_t^2}\right] = o(1/t) . \tag{3}$$

Note that if $Y_0, Y_1, Y_2, \dots$ are not random, but rather chosen deterministically, assumption (2) is satisfied trivially, and assumption (3) follows from assumption (1) and can thus be omitted. Furthermore, if $\mathbb{E}\left[Y_t^2\right] = o(t)$, then it can also be shown that assumption (2) and assumption (3) follow noting that $D_t \geq t$ and $\mathbb{E}[|D_t - \mathbb{E}[D_t]|] \leq \sqrt{VAR[D_t]}$. Letting $m_{k,t}$ denote the number of nodes of degree k at time t (i.e., in the hypergraph H_t), and n_t denote the total number of nodes at time t, we note that

$$m_{k,t} \leq \sum_j m_{j,t} = n_t \leq O(t) \text{ for every } k \geq 1 . \tag{4}$$

Main results. The main result of the paper is the following theorem.

Theorem 1. *In the $H(H_0, p, Y)$ model, assuming Eq.* (1), (2) *and* (3)*, and letting*

$$M_k \; = \; \frac{\Gamma}{1+\Gamma} \cdot \prod_{j=1}^{k-1} \frac{j}{j+1+\Gamma} \; , \tag{5}$$

for every fixed $k \geq 1$ we have

$$\lim_{t\to\infty} \frac{\mathbb{E}\left[m_{k,t}\right]}{p \cdot t} \; = \; M_k \; . \tag{6}$$

Furthermore, if Y_t is deterministic or if $\mathbb{E}\left[Y_t^2\right] = o(t)$, then assumptions (2) *and* (3) *can be omitted and assumption* (1) *simplifies to* $\lim_{t\to\infty} \frac{D_t}{t(Y_t - p)} \; = \; \Gamma \in (0, \infty)$.

Since M_k is proportional to $k^{-(1+\Gamma)}$ (See. [16]), and the number of nodes at time t is concentrated around $p \cdot t$, the following is a direct corollary of Thm. 1.

Theorem 2. *Under the assumptions of Thm. 1, the expected degree distribution of a hypergraph $H(H_0, p, Y)$ follows a power law with exponent $\beta = 1 + \Gamma$. I.e., as $t \to \infty$, the expected fraction of nodes with degree k is proportional to $k^{-\beta}$.*

3 Discussion and Examples

Social network models have contributed much to our understanding of human society in many fields, such as knowledge retrieval, marketing, economics, sociology, and more. However, social structures are complicated and cannot be fully modeled as ordinary networks (or graphs). They are composed of dyads, triads, families, clans, tribes, and communities. Other than dyads, all of these structures are more accurately modeled as hyperedges, rather than ordinary edges. For example, many social network activities, such as coauthorship, Instagram, WhatsApp group messages, group photos, and more, involve multi-party groups that are best represented as hyperedges, or - collectively - as a hypergraph. However, theory has not yet caught up with this type of data organization. By introducing and analyzing a model of preferential attachment hypergraphs, the current paper attempts to narrow this gap. We next discuss some of the implications of our result and provide some examples from simulations and real data.

Result Validation. We first validate our theoretical results via simulations. Recall that our results hold with expectation and not necessarily with high probability. Figure 3 presents the cumulative degree distribution of two hypergraphs. The first hypergraph is a 7-uniform hypergraph (i.e., all hyperedges are of size 7) and with a vertex arrival probability of $p = 0.5$. From Thm. 2, we have $\beta = 1 + \frac{7}{7-0.5} = 2.07$. We plot the cumulative degree distribution with the fit that follows from the theoretical probability of Eq. (5) which lead to a power law degree distribution with exponent β. We repeat the same process for a hypergraph with two edge sizes, 2 or 3, with equal probability. I.e. $Y_t = 2$ or $Y_t = 3$, each with probability 0.5, and a vertex arrival probability of $p = 0.5$. For this case, we have $\beta = 1 + \frac{2.5}{2.5-0.5} = 2.25$. Again we compare the resulting distribution with the theoretical results. For both examples, the theoretical results provide a good fit.

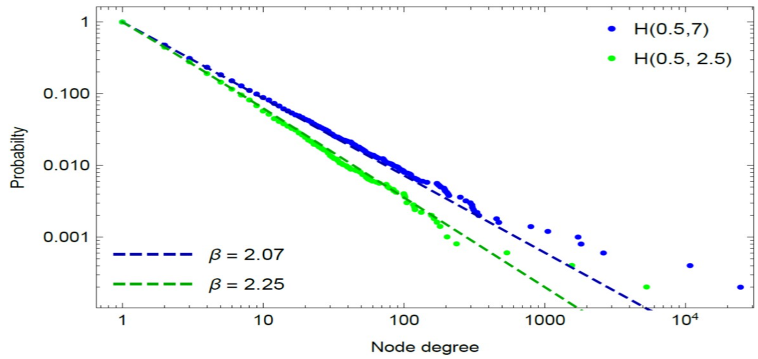

Fig. 3 The cumulative degree distributions of two example hypergraphs with the corresponding β as computed in the paper. $H(0.5,7)$ is a 7-uniform hypergraph with hyperedge size 7 and an edge arrival event probability of 0.5. Theorem 2 implies that $\beta = 1 + \frac{7}{7-0.5} = 2.07$. $H(0.5,2.5)$ is a hypergraph with hyperedge sizes of 2 or 3 with equal probability and an edge arrival event probability of 0.5. Theorem 2 yields $\beta = 1 + \frac{2.5}{2.5-0.5} = 2.25$.

Observed Graphs. Hypergraph-structured social organizations are often *approximated* by graphs. One typical concrete way of achieving this is by replacing each hyperedge e of the given hypergraph H with a collection of graph edges forming a clique subgraph $C(e)$ over the vertices of e (see Figure 1). The resulting social network, denoted $G(H)$, is sometimes referred to as the *observed graph* of the original hypergraph H. A natural question that arises when considering this approach, however, concerns the accuracy of such approximation. What is the price of simplifying the given social structures into graphs?

To address this question, we compare the properties of generated observed graphs against those of preferential attachment hypergraphs. The comparison is based on analyzing the resulting degree distribution and power law parameters in the two representations. Our results suggest that while approximations of hypergraphs by observed graphs may be qualitatively correct, they are typically quantitatively inaccurate. While both the preferential attachment hypergraph model and the observed graph approximation model exhibit a power law degree distribution, the *exponent* of the distribution *changes* between the two models.

We first consider the special case of $H(H_0, p, Y)$ with constant hyperedge size $Y_t = d$. For simplicity, we denote the resulting *d-uniform* hypergraph by $H(p,d)$.

Corollary 3.1 *The degree distribution of a d-uniform hypergraph $H(p,d)$ follows a power law with $\beta(H) = 1 + d/(d-p)$.*

As mentioned earlier, in many cases it is only possible to access the observed graph $G(H)$ that results from the underlying hypergrph H. (Recall that the set of nodes of $G(H)$ is identical to that of H, and for every hyperedge $e \in H$, edges are created in $G(H)$ to form a clique $C(E)$ between all the nodes in e.) For this case one can now prove the following:

Claim 3.2 *The degree distribution of the observed graph $G(H(p,d))$ resulting from a d-uniform hypergraph $H(p,d)$ follows a power law with $\beta(G) = 1 + d/(d-p)$.*

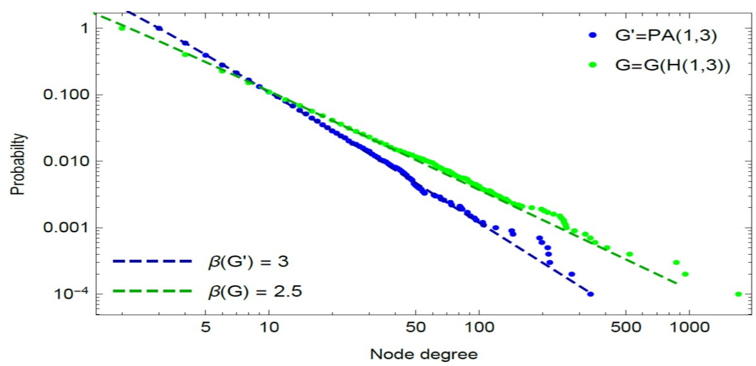

Fig. 4 The cumulative degree distributions of two example networks: an observed graph $G = G(H(1,3))$ derived from a 3-uniform hypergraph $H(1,3)$ with average degree 6 (in green), and a preferential attachment graph $G' = PA(1,3)$ with average degree 6, same as $G(H(1,3))$ (in blue), and their corresponding theoretical β values as derived from our results. Note that $\beta(G) \neq \beta(G')$.

This above claim is supported by simulation results depicted in Figure 4. We show the cumulative degree distribution of the observed graph $G = G(H(1,3))$ where $H(1,3)$ is a 3-uniform hypergraph. By Claim 3.2, $\beta(H(1,3)) = 1 + \frac{3}{3-1} = 2.5$. As can be seen in the figure, $\beta(H(1,3))$ is a good fit to the cumulative degree distribution of G, so $\beta(G) = \beta(H)$.

Note that G has an average degree of 6, since at each time step, one hyperedge of size 3 is added to the network, which gets translated into 3 graph edges in G. We therefore pose the following interesting question. Suppose we generate such a new graph $G' = PA(1,3)$, with expected degree 6, according to the classical graph preferential attachment model. Then each new node must join G' with 3 new edges, just as in G. But what will be the β value of the degree distribution of G' in this case? Chung and Lu's result for preferential attachment graphs [4] implies that $\beta(G') = 1 + 2/(2-1) = 3$. Hence $\beta(G)$ and $\beta(G')$, the *observed* degree distributions of G and G' respectively, turn out to be different (see Figure 4).

An immediate implication of this example is that one should be careful in determining the appropriate model to capture the observed degree distribution for a given application. In particular, one must decide if the generative model is that of a hypergraph or that of the classical graph model.

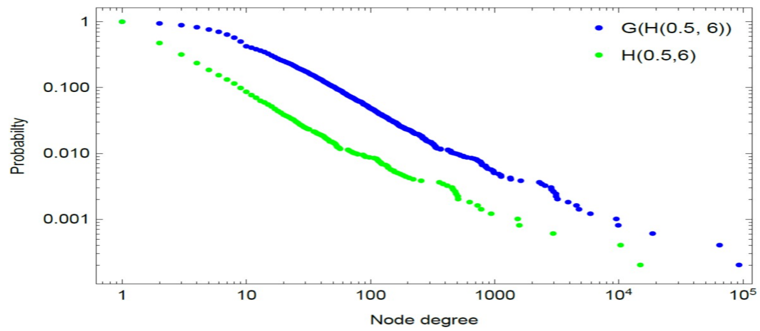

Fig. 5 The cumulative degree distributions of a hypergraph $H(0.5,6)$ and its observed graph $G(H(0.5,6))$. H has hyperedges of expected size 6. However, it is a *nonuniform* hypergraph, i.e., hyperedge sizes are selected at random between sizes 2 to 9. The edge arrival event probability is 0.5. The two cumulative degree distributions are different, particularly on low degrees.

The situation is even more complex for nonuniform hypergraphs, as here, Claim 3.2 no longer holds. Figure 5 presents the cumulative degree distribution of a hypergraph $H(0.5,6)$ and its observed graph $G(H(0.5,6))$. The hypergraph has an expected hyperedge of size 6. However, it is a *nonuniform* hypergraph, i.e., hyperedge sizes are selected uniformly at random between sizes 2 to 9. The edge arrival event probability is 0.5. One can clearly see that the two cumulative degree distributions are different, in particular on low degrees. The hypergraph cumulative degree distribution behaves linearly, while the cumulative degree distribution of the observed graph is more complex and is curved in the low degrees. It follows that choosing to model a social network as either a graph or a hypergraph will influence the analysis. Stated more explicitly, when the dataset comes from a hypergraph, using a hypergraph model instead of an observed graph model is clearly a better choice.

Figure 5 reveals another interesting connection in network theory and complex systems. According to Newman's seminal work [5], when calculating the exponent of the power law, it is a good practice to neglect all the nodes whose degree is among the (roughly) 6 lowest degrees in the network. This is a very useful idea which, in fact, works very well in practice. However, deleting these smallest degree nodes can, in certain cases, be akin to ignoring 90% of the data in a network. The hypergraph model suggested herein can be used to explain *why* the lowest-degree nodes are such "troublemakers": this is because these nodes come from hypergraphs. Indeed, when moving from a graph to a hypergraph, the lowest-degree nodes *do* suddenly behave well, as can be seen in Figure 5.

Real Data, Hypergraph Example. In this section we consider a real hypergraph, the coauhtorship social network `Co-HEPT`, constructed from research papers that were published in the High Energy Physics Theory section of arXiv in the years 1991-2003 and were part of KDD cup 2003 [9]. In `Co-HEPT`, each author is a vertex and each research paper is a hyperedge connecting the authors of the paper.

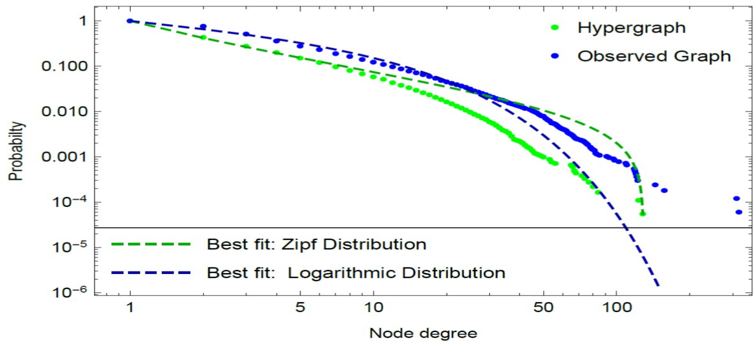

Fig. 6 An example of the hypergraph of the coauthorship network `Co-HEPT`. Every paper is a hyperedge. The cumulative degree distribution of both the hypergraph and the observed graph are plotted together with the best fit for the distributions. For the hypergraph we get a power law distribution (Zipf with cutoff), but the observed graph exhibits a logarithmic distribution.

Over all there are 18,268 authors and 25,168 hyperedges (i.e., papers). The edge sizes range from 1 to 8, with an average of 2.51 authors per paper.

Figure 6 shows the cumulative degree distribution of both the hypergraph (green) and the observed graph (blue) of this network. To demonstrate our discussion above we used Mathematica to find the best fit distribution for the degree distribution (note that we show the cumulative degree distribution, but the fitting was done directly on the degree distribution [5]). As we claim the result are different, while the best fit for the hypergraph data (darker dashed green) is a power law degree distribution (a Zipf distribution with cutoff) the best fit for the observed graph (darker dashed blue) was a logarithmic distribution. We compared the root mean square error of both distributions with the real data of the hypergraph and the Zipf distribution produced an error which was 3.5 times smaller than the error of the logarithmic distribution. Moreover, when we enforced the best fit of the observed graph to be a power law, the results were even worse with different beta.

4 Proof of the Main Theorem

We now prove the main result of the paper, Theorem 1. During our analysis we use the following lemma, which can be found in [4].

Lemma 1. *[4] Suppose that a sequence $\{a_t\}$ satisfies the recurrence relation*

$$a_{t+1} = \left(1 - \frac{b_t}{t}\right) a_t + c_t \quad \text{for} \quad t \geq t_0 \,.$$

Furthermore, suppose $\lim_{t\to\infty} b_t = b > 0$ and $\lim_{t\to\infty} c_t = c$. Then $\lim_{t\to\infty} a_t/t$ exists and $\lim_{t\to\infty} a_t/t = c/(1+b)$.

As a first step, we find a recursive representation for $m_{k,t+1}$ in terms of $m_{k',t}$ for smaller $k' \leq k$. Note that $m_{0,t} = m_{0,0}$ for every t. This follows nodes of degree 0 cannot be included in future edges, and every new node arrives with degree 1. Therefore, we focus on $m_{k,t}$ for $k \geq 1$. Let $\mathscr{F}_t$ denote the σ-algebra generated by the hypergraphs $H_0, \ldots, H_t$. Intuitively, $\mathscr{F}_t$ encodes all the history until time t. For $k \geq 2$ we have

$$\mathbb{E}\left[m_{k,t+1} \mid \mathscr{F}_t, Y_t\right] \;=\; \sum_v \mathbb{P}[d_{t+1}(v) = k \mid \mathscr{F}_t, Y_t] \;=\; \sum_{j=0}^{\min\{Y_t, k-1\}} m_{k-j,t}\ell^t_{k-j,k} \;,$$

where $\ell^t_{k-j,k}$ is the probability that a node with degree $k-j$ at time t will have degree k at time $t+1$. Note that $\ell^t_{k-j,k} = 0$ for every $j > Y_t$ since at time t the degree could increase by at most the size of the edge added at time t, namely Y_t. Also note that $m_{k-j,t} = 0$ for every $j \geq k$ since all nodes are of degree at least 1. In order to evaluate $\ell^t_{k-j,k}$, we denote $\mathbb{P}(i;n,q) = \binom{n}{i} \cdot q^i (1-q)^{n-i}$, and consider the following two possible cases for a node v satisfying $d_t(v) = k-j$ and $d_{t+1}(v) = k$:

1. $0 \leq j \leq \min\{Y_t, k-1\}$, at time t no new nodes were added, and the new edge contained v exactly j times (this happens with probability $(1-p) \cdot \mathbb{P}(j; Y_t, (k-j)/D_t)$)
2. $0 \leq j \leq \min\{Y_t - 1, k-1\}$, at time t a new node was added and the new edge contained v exactly j times (this happens with probability $p \cdot \mathbb{P}(j; Y_{t-1}, (k-j)/D_t)$)

Thus, for every $k \geq 2$ we have

$$\begin{aligned}\mathbb{E}\left[m_{k,t+1} \mid \mathscr{F}_t, Y_t\right] \;=\; &(1-p) \sum_{j=0}^{\min\{Y_t, k-1\}} m_{k-j,t} \cdot \mathbb{P}\left(j; Y_t, \frac{k-j}{D_t}\right) \\ &+\; p \sum_{j=0}^{\min\{Y_t-1, k-1\}} m_{k-j,t} \cdot \mathbb{P}\left(j; Y_t - 1, \frac{k-j}{D_t}\right) \;. \qquad (7)\end{aligned}$$

Similarly, for $k = 1$ we have

$$\mathbb{E}\left[m_{1,t+1} \mid \mathscr{F}_t, Y_t\right] \;=\; (1-p) \cdot m_{1,t} \cdot \left(1 - \frac{1}{D_t}\right)^{Y_t} + p \cdot m_{1,t} \cdot \left(1 - \frac{1}{D_t}\right)^{Y_t - 1} + p \;, \quad (8)$$

where here we added a p to account for the new node that arrives with probability p. We now show that the resulting network follows a power law. To this end, we note that combining Ineq. (4) and Eq. (2), the following lemma follows.

Lemma 2. *For every $t \geq 0$ and fixed $k \geq 0$, we have*

$$\left| \mathbb{E}\left[\frac{m_{k,t} \cdot k}{D_t}\right] - \frac{\mathbb{E}\left[m_{k,t} \cdot k\right]}{\mathbb{E}\left[D_t\right]} \right| \;\leq\; o(1) \;,$$

where $o(1)$ tends to 0 as $t \to \infty$.

We now prove Theorem 1.

Proof of Theorem 1: We prove Eq. (6) by induction on k. For $k = 1$, taking expectation on both sides of Eq. (8) yields

$$\mathbb{E}[m_{1,t+1}] = \mathbb{E}\left[m_{1,t} \cdot \left(1 - \frac{1}{D_t}\right)^{Y_t - 1} \left(1 - \frac{1-p}{D_t}\right)\right] + p\ .$$

Hence

$$\begin{aligned}\mathbb{E}[m_{1,t+1}] \ &\geq\ \mathbb{E}\left[m_{1,t} \cdot \left(1 - \frac{Y_t - 1}{D_t}\right)\left(1 - \frac{1-p}{D_t}\right)\right] + p \\ &\geq\ \mathbb{E}\left[m_{1,t} \cdot \left(1 - \frac{Y_t - p}{D_t}\right)\right] + p\ \geq\ \mathbb{E}[m_{1,t}]\left(1 - \frac{\mathbb{E}[Y_t] - p}{\mathbb{E}[D_t]}\right) + p - o(1)\ ,\end{aligned} \tag{9}$$

where the last inequality follows from Lemma 2 and the independency of Y_t from D_t and $m_{k,t}$. In the other direction we have

$$\begin{aligned}\mathbb{E}[m_{1,t+1}] &\leq \mathbb{E}\left[m_{1,t} \cdot \frac{1}{1 + (Y_t - 1)/D_t}\left(1 - \frac{1-p}{D_t}\right)\right] + p \\ &= \mathbb{E}\left[m_{1,t} \cdot \left(1 - \frac{Y_t - p}{D_t + Y_t - 1}\right)\right] + p\ \leq\ \mathbb{E}\left[m_{1,t} \cdot \left(1 - \frac{Y_t - p}{D_t} + \frac{Y_t^2}{D_t^2}\right)\right] + p \\ &\leq \mathbb{E}\left[m_{1,t} \cdot \left(1 - \frac{Y_t - p}{D_t}\right)\right] + \mathbb{E}\left[O(t) \cdot \frac{Y_t^2}{D_t^2}\right] + p \\ &\leq \mathbb{E}[m_{1,t}]\left(1 - \frac{\mathbb{E}[Y_t] - p}{\mathbb{E}[D_t]}\right) + p + o(1)\ ,\end{aligned} \tag{10}$$

where the penultimate inequality follows from Ineq. (4), and the last inequality follows from Eq. (3) and Lemma 2. Considering Ineq. (9) and (10), and applying Lemma 1 with (a_t, b_t, c_t) set to $(\mathbb{E}[m_{1,t}], (\mathbb{E}[Y_t] - p)t/\mathbb{E}[D_t], p \pm o(1))$, with the corresponding limits $b = 1/\Gamma$ and $c = p$ obtained by Eq. (1), we have

$$\lim_{t\to\infty} \frac{\mathbb{E}[m_{1,t}]}{t}\ =\ \frac{p}{1 + 1/\Gamma}\ =\ p \cdot M_1\ .$$

Hence the induction basis follows. Next, assuming the claim for k, we prove it for $k+1$. Taking expectation on both sides of Eq. (7), we write

$$\mathbb{E}\left[m_{k,t+1}\right] = \Psi + (1-p) \cdot \mathbb{E}[\varphi(Y_t)] + p \cdot \mathbb{E}[\varphi(Y_t - 1)]\ , \tag{11}$$

where Ψ represents the expectation of the first two terms in Eq. (7), namely,

$$\begin{aligned}\Psi &= (1-p)\sum_{j=0}^{1}\mathbb{E}\left[m_{k-j,t}\cdot\mathbb{P}\left(j;Y_t,\frac{k-j}{D_t}\right)\right]+p\sum_{j=0}^{1}\mathbb{E}\left[m_{k-j,t}\cdot\mathbb{P}\left(j;Y_t-1,\frac{k-j}{D_t}\right)\right]\\ &= \mathbb{E}\left[m_{k,t}\left((1-p)\cdot\left(1-\frac{k}{D_t}\right)^{Y_t}+p\cdot\left(1-\frac{k}{D_t}\right)^{Y_t-1}\right)\right]\\ &+ \mathbb{E}\left[m_{k-1,t}\cdot\frac{k-1}{D_t}\left(Y_t(1-p)\left(1-\frac{k-1}{D_t}\right)^{Y_t-1}+p(Y_t-1)\left(1-\frac{k-1}{D_t}\right)^{Y_t-2}\right)\right] .\end{aligned}$$

and the expectation of the remaining terms is represented using the notation

$$\varphi(n) \;=\; \sum_{j=2}^{\min\{n,k-1\}} m_{k-j,t}\cdot\mathbb{P}\left(j;n,\frac{k-j}{D_t}\right) .$$

Considering Eq. (11), we show that Ψ is the dominant term and both $\mathbb{E}\left[\varphi(Y_t)\right]$ and $\mathbb{E}\left[\varphi(Y_t-1)\right]$ are negligible. By Ineq. (4), we have

$$\begin{aligned}\varphi(Y_t) &= \sum_{j=2}^{\min\{Y_t,k-1\}} m_{k-j,t}\cdot\binom{Y_t}{j}\cdot\left(\frac{k-j}{D_t}\right)^{j}\cdot\left(1-\frac{k-j}{D_t}\right)^{Y_t-j}\\ &\le O(t)\sum_{j=2}^{\min\{Y_t,k-1\}}\binom{Y_t}{j}\cdot\left(\frac{k-j}{D_t}\right)^{j}\cdot\left(1-\frac{k-j}{D_t}\right)^{Y_t-j}\\ &\le O(t)\sum_{j=2}^{\min\{Y_t,k-1\}} Y_t^j\cdot\left(\frac{k}{D_t}\right)^{j}\cdot\left(1-\frac{1}{D_t}\right)^{Y_t-k+1}\\ &\le O(t)\cdot\frac{Y_t^2\cdot k^2}{D_t^2}\cdot e^{-Y_t/D_t}\sum_{j=2}^{\min\{Y_t,k-1\}}\left(\frac{Y_t\cdot k}{D_t}\right)^{j-2} .\end{aligned}$$

If $Y_t \le D_t$, we have

$$\varphi(Y_t) \le O(t)\cdot\frac{Y_t^2\cdot k^2}{D_t^2}\cdot k^{k-2} \;=\; O\left(t\cdot\frac{Y_t^2}{D_t^2}\right) ,$$

whereas if $Y_t \ge D_t$, we have

$$\begin{aligned}\varphi(Y_t) &\le O(t)\cdot\frac{Y_t^2\cdot k^2}{D_t^2}\cdot e^{-Y_t/D_t}\cdot\frac{(Y_t\cdot k/D_t)^{k-2}-1}{(Y_t\cdot k/D_t)-1}\\ &\le O(t)\cdot\frac{Y_t^2\cdot k^k}{D_t^2}\cdot e^{-Y_t/D_t}\cdot\frac{(Y_t/D_t)^{k-2}}{k-1}\\ &\le O(t)\cdot\frac{Y_t^2\cdot k^k}{D_t^2}\cdot e^{-(k-2)}\cdot\frac{(k-2)^{k-2}}{k-1} \;\le\; O\left(t\cdot\frac{Y_t^2}{D_t^2}\right) ,\end{aligned}$$

where the penultimate inequality follows since the expression $e^{-x}\cdot x^{\alpha}$ is maximized at $x=\alpha$. Thus in any case, we obtain

$$\varphi(Y_t) \le O\left(t \cdot \frac{Y_t^2}{D_t^2}\right) .$$

Therefore, by Eq. (3),

$$\mathbb{E}\left[\varphi(Y_t)\right] \;\le\; o(1) . \tag{12}$$

Similarly

$$\mathbb{E}\left[\varphi(Y_t - 1)\right] \;\le\; o(1) . \tag{13}$$

We now provide tight bounds for Ψ. Similar to the case where $k = 1$, we have

$$\begin{aligned}\Psi \;&\le\; \mathbb{E}\left[m_{k,t} \cdot \left(1 - \frac{(Y_t - p)k}{D_t}\right)\right] + \mathbb{E}\left[O(t) \cdot \frac{Y_t^2}{D_t^2}\right] \cdot k \\ &+\; \mathbb{E}\left[m_{k-1,t} \cdot \frac{k-1}{D_t} \cdot \left((1-p) \cdot Y_t + p \cdot (Y_t - 1)\right)\right] \\ &\le\; \mathbb{E}\left[m_{k,t}\right]\left(1 - \frac{k(\mathbb{E}\left[Y_t\right] - p)}{\mathbb{E}\left[D_t\right]}\right) \;+\; \mathbb{E}\left[m_{k-1,t}\right] \cdot \frac{k-1}{\mathbb{E}\left[D_t\right]} \cdot \left(\mathbb{E}\left[Y_t\right] - p\right) + o(1) ,\end{aligned} \tag{14}$$

where the last inequality follows from Eq. (3) and Lemma 2. We now derive a lower bound for Ψ. We have

$$\begin{aligned}\Psi &\ge \mathbb{E}\left[m_{k,t}\left((1-p)\left(1 - \frac{k \cdot Y_t}{D_t}\right) + p\left(1 - \frac{k(Y_t - 1)}{D_t}\right)\right)\right] \\ &+ \mathbb{E}\left[m_{k-1,t} \cdot \frac{k-1}{D_t} \cdot \left(1 - \frac{(k-1)(Y_t - 1)}{D_t}\right) \cdot \left(Y_t(1-p) + p(Y_t - 1)\right)\right] \\ &= \mathbb{E}\left[m_{k,t}\left(1 - \frac{k(Y_t - p)}{D_t}\right)\right] + \mathbb{E}\left[m_{k-1,t} \cdot \frac{k-1}{D_t} \cdot \left(1 - \frac{(k-1)(Y_t - 1)}{D_t}\right)(Y_t - p)\right] .\end{aligned}$$

By Lemma 2 and Eq. (4), we have

$$\begin{aligned}\Psi \;&\ge\; \mathbb{E}\left[m_{k,t}\right]\left(1 - \frac{k(\mathbb{E}\left[Y_t\right] - p)}{\mathbb{E}\left[D_t\right]}\right) - o(1) \\ &+\; \mathbb{E}\left[m_{k-1,t}\right] \cdot \frac{k-1}{\mathbb{E}\left[D_t\right]} \cdot \left(\mathbb{E}\left[Y_t\right] - p\right) - \mathbb{E}\left[O\left(t \cdot \frac{Y_t^2}{D_t^2}\right)\right] \\ &=\; \mathbb{E}\left[m_{k,t}\right]\left(1 - \frac{k(\mathbb{E}\left[Y_t\right] - p)}{\mathbb{E}\left[D_t\right]}\right) + \mathbb{E}\left[m_{k-1,t}\right] \cdot \frac{k-1}{\mathbb{E}\left[D_t\right]} \cdot \left(\mathbb{E}\left[Y_t\right] - p\right) - o(1) ,\end{aligned} \tag{15}$$

where the last inequality follows from Ineq. (3). Plugging Ineq. (12), (13), (14) and (15) into Eq. (11), we obtain

$$\mathbb{E}\left[m_{k,t+1}\right] \;=\; \mathbb{E}\left[m_{k,t}\right] \cdot \left(1 - \frac{(\mathbb{E}\left[Y_t\right] - p)k}{\mathbb{E}\left[D_t\right]}\right) + \mathbb{E}\left[m_{k-1,t}\right] \cdot \frac{k-1}{\mathbb{E}\left[D_t\right]} \cdot \left(\mathbb{E}\left[Y_t\right] - p\right) \pm o(1) . \tag{16}$$

Therefore, applying Lemma 1 with (a_t, b_t, c_t) set to

$$\left(\mathbb{E}\left[m_{k,t}\right]\ ,\ \frac{(\mathbb{E}\left[Y_t\right]-p)kt}{\mathbb{E}\left[D_t\right]}\ ,\ \mathbb{E}\left[m_{k-1,t}\right]\cdot\frac{k-1}{\mathbb{E}\left[D_t\right]}\cdot(\mathbb{E}\left[Y_t\right]-p)\pm o(1)\right)$$

with the corresponding limits $b = k/\Gamma$ and $c = (\lim_{t\to\infty}\mathbb{E}\left[m_{k-1,t}\right]/t)\cdot(k-1)/\Gamma$ obtained by Eq. (1), we have

$$\lim_{t\to\infty}\frac{\mathbb{E}\left[m_{k,t}\right]}{t} = \lim_{t\to\infty}\frac{\mathbb{E}\left[m_{k-1,t}\right]}{t}\cdot\frac{(k-1)/\Gamma}{1+k/\Gamma}\ .$$

Hence, by the induction hypothesis, we obtain

$$M_k = M_{k-1}\cdot\frac{k-1}{k+\Gamma}\ ,$$

thus concluding the proof of Thm. 1. □

5 Conclusion

We conclude with some discussion and directions for future work. In the future, hypergraph models are likely to replace graph models at the core of social structure analysis. This is due to the fact that hypergraph data is available, and hypergraphs generate considerably more accurate pictures of the reality of social structures, as this paper shows. It is therefore hoped that the preferential attachment hypergraph model, as suggested herein, will contribute a modest step towards the next leap in the study of social structures.

It may be interesting to study next a setting where p is not constant but rather a sequence, p_t, which depends on the time t and tends to 0, i.e., $\lim_{t\to\infty} p_t = 0$. We find this to be an exciting new research direction, which can provide explanations to a variety of phenomena that are not captured in the current model.

References

1. C. Avin, Y. Lando and Z. Lotker. Radio cover time in hyper-graphs. *Ad Hoc Networks*, 12:278–290, 2014.
2. A-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
3. B. Bollobás and P. Erdös. Cliques in random graphs. In *Mathematical Proc. Cambridge Philosophical Soc.*, volume 80, pages 419–427. Cambridge Univ Press, 1976.
4. F.R.K. Chung and L. Lu. *Complex graphs and networks*. AMS, 2006.
5. A. Clauset, C. R. Shalizi and M. E.J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.
6. C. Cooper, A. Frieze, M. Molloy and B. Reed. Perfect matchings in random r-regular, s-uniform hypergraphs. *Combinatorics, Probability and Computing*, 5:1–14, 1996.
7. D. Ellis and N. Linial. On regular hypergraphs of high girth. *Prepr. arXiv:1302.5090*, 2013.
8. P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*, 5:17–61, 1960.
9. J. Gehrke, P. Ginsparg and J. Kleinberg. Overview of the 2003 KDD Cup. *SIGKDD Explorations*, 5(2):149–151, 2003.
10. G. Ghoshal, V. Zlatić, G. Caldarelli and M. E.J. Newman. Random hypergraphs and their applications. *Physical Review E*, 79(6):066118, 2009.
11. S. Lattanzi and D. Sivakumar. Affiliation networks. In *Proc. 41st ACM STOC*, 427–434, 2009.
12. M. E.J. Newman. *Networks: An Introduction*. Oxford Univ. Press, 2010.
13. M. E.J. Newman, D. J. Watts and S. H. Strogatz. Random graph models of social networks. *Proc. Nat. Acad. Sci.*, 99(suppl 1):2566–2572, 2002.
14. D. de Solla Price. A general theory of bibliometric and other cumulative advantage processes. *J. Amer. Soc. Inform. Sci.*, 27(5):292–306, 1976.
15. J.-W. Wang, L.-L. Rong, Q.-H. Deng and J.-Y. Zhang. Evolving hypernetwork model. *The European Physical Journal B*, 77(4):493–498, 2010.
16. J. G. Wendel. Note on the gamma function. *The Amer. Math. Monthly*, 55:563–564, 1948.
17. Z.-K. Zhang and C. Liu. A hypergraph model of social tagging networks. *J. Statistical Mechanics: Theory and Experiment*, 2010(10):P10005, 2010.