Internet Mathematics Vol. 8, Nos. 1–2: 113–160

Multiplicative Attribute Graph Model of Real-World Networks

Myunghwan Kim and Jure Leskovec

Abstract. Networks are a powerful way to describe and represent social, technological, and biological systems, where nodes represent entities (people, web sites, genes) and edges represent interactions (friendships, communication, regulation). The study of such networks then seeks to find common structural patterns and explain their emergence through tractable models of network formation.

In most networks, each node is associated with a rich set of attributes or features. For example, users in online social networks have profile information, genes have properties and functions, and web pages contain text. However, most existing network models focus on modeling the network structure while ignoring the features and properties of the nodes. Thus, the questions that we address in this work are as follows: What is a mathematically tractable model that naturally captures ways in which the network structure and node attributes interact? What are the properties of networks that arise under such a model?

We present a model of network structure that we refer to as the *multiplicative attribute graphs (MAG)* model. The MAG model naturally captures the interactions between the network structure and the node attributes. We consider a model in which each node has a vector of categorical attributes associated with it. The link-affinity matrix then models the interaction between the value of a particular attribute and the probability of a link between a pair of nodes. The MAG model yields itself to mathematical analysis, and we derive thresholds for the connectivity and the emergence of the giant connected component, and show that the model gives rise to networks with a constant diameter. We also analyze the degree distribution and find surprising flexibility of the MAG model in that it can generate networks with either log-normal or power-law degree distribution.

ISSN: 1542-7951 print

1. Introduction

Networks have emerged as a main tool for studying phenomena across the social, technological, and natural worlds. And with the emergence of the Internet, social media, and high-throughput gene expression analysis, massive amounts of network data have become available. Such networks have been thoroughly studied, and a unifying theme of studying real-world networks is to find patterns of connectivity and explain them through models. The goal is to find answers to questions such as, "What do real graphs look like?" "How do they evolve over time? "How can we synthesize realistic-looking graphs?" "How can we find models that explain the observed patterns?" and "What are algorithmic consequences of the observations and models?"

Research on networks consists in empirical observations about the structure of networks and the models giving rise to such structures. The empirical analysis of networks aims to discover common structural properties or patterns [Chakrabarti and Faloutsos 06], such as heavy-tailed degree distributions [Faloutsos et al. 99, Broder et al. 00], local clustering of edges [Watts and Strogatz 98, Leskovec et al. 09], small diameters [Albert et al. 99, Leskovec et al. 05b], navigability [Milgram 67, Kleinberg 00], emergence of network community structure [Fortunato 10, Girvan and Newman 02, Leskovec et al. 08], and so on. In parallel, there have been efforts to develop the network-formation mechanisms that naturally generate networks with the observed structural features. In these network-formation mechanisms, there have been two relatively dichotomous modeling approaches.

Broadly speaking, one line of work has focused mainly on relatively simple "mechanistic" but mathematically tractable network models in which connectivity patterns observed in the real world naturally emerge from the model. The prime example in this line of research is the preferential attachment model with its many variants [Barabási and Albert 99, Aiello et al. 00, Bollobás and Riordan 03, Borgs et al. 07, Cooper and Frieze 03]. The model specifies a simple but very natural edge-creation mechanism that in the limit leads to networks with power-law degree distributions. Other models of similar flavor include the copying model [Kumar et al. 00], the small-world model [Watts and Strogatz 98, Kleinberg 00], geometric random graphs [Flaxman et al. 04], the forest fire model [Leskovec et al. 05b], the random surfer model [Blum et al. 06], and models of bipartite affiliation networks [Lattanzi and Sivakumar 09].

On the other hand, a different approach to modeling network data has also emerged. The effort here is in the development of statistically sound models that consider the structure of the network as well as the features of nodes and edges in the network (for example, in a social network, node features could include age, gender, hometown, and profession of a user). Examples of such models include the exponential random graphs [Wasserman and Pattison 96], the stochastic block model [Airoldi et al. 07], and the latent space model [Hoff and Raftery 02]. Such models are generally not analyzed mathematically in a sense that one would prove theorems about the properties of networks that emerge from the model but are rather fit to network data in order to discover interesting facts about a particular network data set.

1.1. "Mechanistic" and "Statistical" Models

Generally, there has been some gap between the above two lines of research. The "mechanistic" models are analytically tractable in a sense that one can mathematically analyze properties of the networks that arise from the models. These models emphasize the natural emergence of networks that have certain structural properties found in real-world networks. However, such models are usually not statistically "interesting" in the sense that they are often too simplistic to accurately model the heterogeneities in linking behavior of different nodes. On the other hand, "statistical" models are usually analytically intractable, since they do not lend themselves to mathematical analysis. Even though the number of parameters in such models is usually large, the models come accompanied with statistical procedures for model parameter estimation. Such models have proven to be very useful for testing various hypotheses about the interaction of network structure and the attributes of nodes and edges.

Although models of network structure and formation are seldom both analytically tractable and statistically interesting, an example of a model satisfying both features is the Kronecker graphs model [Leskovec et al. 05a, Leskovec et al. 10, Weichsel 62], which is based on the recursive tensor product of small graph adjacency matrices. The model is analytically tractable in a sense that one can analyze global structural properties of networks that emerge from the model [Bodine et al. 10, Leskovec et al. 10, Mahdian and Xu 07]. In addition, this model is statistically meaningful because there exists an efficient parameter-estimation technique based on maximum likelihood [Gleich and Owen 11, Kim and Leskovec 11b, Leskovec and Faloutsos 07]. It has been empirically shown that with only four parameters, Kronecker graphs quite accurately model the global structural properties of real-world networks such as degree distributions, edge clustering, diameter, and spectral properties of the graph adjacency matrices. However, even though the Kronecker graphs model is able to capture the global structure of networks, the model considers only "bare" networks whose nodes have no attributes.

1.2. Modeling Networks with Node Attribute Information

Network data usually contain not only the node connectivity information but also attributes and features of the nodes. For example, in social networks we are given a list of friends of a person as well as his or her attributes, such as gender, workplace, and hobbies. When studying biological networks, we are given the connectivity information as well as the properties of genes or proteins that constitute the nodes of the network.

In order to accurately model networks, node characteristics as well as the network connectivity structure need to be considered simultaneously. However, the attempt to model the interaction between the network structure and node attributes raises a wide range of questions. For instance, how do we account for the heterogeneity in the population of nodes, and how do we combine node features in a natural way to obtain probabilities of individual links? While earlier work on a general class of latent space models [Young and Scheinerman 07, Hoff and Raftery 02] attempted to address such questions, most resulting models were either analytically tractable but statistically uninteresting or statistically very powerful but hard to analyze mathematically.

To bridge this gap, we propose a class of stochastic network models that we refer to as *multiplicative attribute graphs (MAG)*. The model naturally captures the interactions between occurrence of links in the network and the node attributes in a clean and tractable manner. We consider a model in which each node has a vector of categorical attributes associated with it. Attribute values of a node are then combined in order to model the emergence of links. The model allows for rich interaction between node attributes in a sense that one can simultaneously model attributes that reflect homophily (love of the same) as well as heterophily (love of the different) [Rogers and Bhowmik 70, McPherson et al. 01]. Homophily occurs when people that share a certain feature (that is, people who are similar) are more likely to create links among themselves. On the other hand, heterophily occurs when people are more likely to create connections with those who do not share some feature with them (for example, gender). Our MAG model is designed to capture, in a natural way, homophily, heterophily, and other types of node attribute interactions that occur in networks.

The rest of the paper is organized as follows. We proceed by formulating the model in Section 2. In the sections that follow we then present our mathematical results. Section 3 examines the number of edges and shows that our model naturally obeys the densification power law [Leskovec et al. 05b]. Section 4 examines the connectivity of the MAG model, which includes the conditions not only when the network contains a giant connected component but also when it becomes connected. Section 5 shows that the diameter of the MAG model remains small even though the number of nodes is large. Section 6 shows that networks emerging from the MAG model have a log-normal degree distribution.

Section 7 describes a more general version of the model that can also capture the power-law degree distribution. We view this as particularly interesting in the light of the current debate on the power-law and the log-normal distributions in empirical data [Mitzenmacher 04, Mitzenmacher 06]. Our results imply that the MAG model is flexible in a sense that networks with very different properties emerge depending on the parameter configuration.

Finally, Section 8 verifies the properties of the MAG model by simulation experiments. The results of the simulations examine how the synthetic network changes depending on the parameters as well as on how similar the network looks to real-world networks.

2. Formulation of the Multiplicative Attribute Graph (MAG) Model

In the following section, we introduce the multiplicative attribute graph (MAG) model. We first formulate a general version of the MAG model and then present a simplified version that we will mathematically analyze throughout the paper. Finally, we also investigate the connection to previous efforts on network modeling.

2.1. General Considerations

On the road to formulating the multiplicative attributes graph model, we first introduce the two essential ingredients of the model: note-attribute vectors and the attribute link-affinity matrices.

First, we consider a setting in which each node u of the network has an *attribute vector* $a(u)$ of k categorical attributes associated with it. For example, one can think that we ask each node of the network a sequence of k yes/no questions, such as, “Are you male?” “Do you like ice cream?” and so on. A sequence of answers of a node to such questions then forms (in this case) a binary vector of length k associated with that node.

A second essential ingredient of our model is to specify a mechanism that generates the probability of an edge between two nodes based on their attribute vectors. As mentioned before, we aim to be able to model both the homophily of some features as well as the heterophily of others. To achieve this, we associate

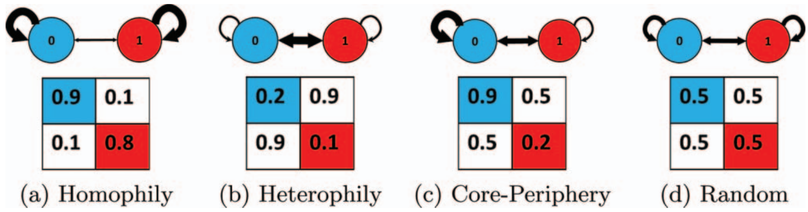


Figure 1. Node-attribute link affinities. Four different types of link affinity of a particular attribute. Circles represent nodes with attribute value 0 and 1, and the width of the arrows corresponds to the affinity of link formation between the two groups. Each bottom figure indicates its corresponding link-affinity matrix (color figure available online).

each attribute i (the ith "question") with the *attribute link-affinity matrix* Θ_i. Each entry of matrix Θ_i captures the affinity of the ith attribute to form a link between a pair of nodes given the value of attribute i for both nodes.

More precisely, $\Theta_i[a_i(u), a_i(v)]$ indicates the affinity with which a pair of nodes u and v form a link, given that each ith attribute of nodes u and v takes value $a_i(u)$ and value $a_i(v)$ respectively. In other words, to obtain the link affinity corresponding to the ith attribute of nodes u and v, the values $(a_i(u), a_i(v))$ of the ith attribute of nodes u and v "select" an appropriate entry of Θ_i. For example, the attribute value of the first node selects the row of matrix Θ_i (row 0 or row 1), and the value of the second node selects the column. Intuitively, the higher the value $\Theta_i[a_i(u), a_i(v)]$, the stronger is the effect of the particular attribute combination $(a_i(u), a_i(v))$ on forming a link.

By defining link-affinity matrices, we can capture the various types of structures in real-world social networks. For example, consider that nodes are described with binary attributes. Then each link-affinity matrix Θ_i is a 2×2 matrix. Figure 1 illustrates four possible link-creation affinities of a single binary attribute, which is denoted by a.

The top row of the figure visualizes the overall structure of the network when one is considering the value of that particular attribute a. Circle 0 represents all the nodes for which attribute a takes the value 0, and circle 1 represents all the nodes with attribute value 1. The width of each arrow indicates the affinity of the link formation between a pair of nodes with a given attribute value. For example, the arrow $0 \longleftrightarrow 1$ indicates the affinity of link formation between a node with 0 value of a given attribute a and a node for which a takes the value 1. Below, we also show the structure of the corresponding attribute link-affinity matrix.

Consider Figure 1(a), which illustrates homophily, which is a tendency of nodes to link with others that have the same value of the particular attribute. This means that nodes sharing the value 0 or 1 are more likely to link than pairs of nodes with different values of the attribute. Such structures can be captured by the link-affinity matrix Θ, which has large values on the diagonal entries, which means that link probability is high when nodes share the same attribute value. The graph at the top of the figure demonstrates that there will be many links between nodes that have the value of the attribute set to 0 and many links between nodes that have the value 1, but there will be few links between nodes where one has the value 0 and the other takes the value 1.

Similarly, Figure 1(b) illustrates heterophily, whereby nodes that do not share the value of the attribute are more likely to link. In the extreme case, such an affinity structure gives rise to bipartite networks.

Furthermore, Figure 1(c) shows the core–periphery affinity [Holme 05, Leskovec et al. 09], whereby links are most likely between the "0 nodes" (members of the core) and least likely between "1 nodes" (members of the periphery). However, links between "zeros" and "ones" are more likely than between the "ones," which means that nodes of the core are the most connected and that the nodes of the periphery are better connected to the core than among themselves [Leskovec 09].

Lastly, Figure 1(d) illustrates the uniform affinity structure, which corresponds to an Erdős–Rényi random graph model, whereby nodes have the same affinity of forming a link regardless of their corresponding attribute values.

These examples indicate that the MAG model provides flexibility in the network structure via link-affinity matrices. Although our examples focused on the simplest case of binary attributes and undirected graphs, the MAG model naturally allows for attributes with higher cardinalities—for the attribute of cardinality d_i, the corresponding link-affinity matrix Θ_i is a $d_i \times d_i$ matrix. Similarly, to model directed graphs, we drop the restriction of Θ_i being symmetric.

2.2. The Multiplicative Attributes Graph (MAG) Model

Now we formulate a general version of the MAG model. To start, let each node $u \in V$ have a vector of k categorical attributes and let each attribute have cardinality d_i for $i = 1, 2, \ldots, k$. We also have k link-affinity matrices $\Theta_i \in d_i \times d_i$ for $i = 1, 2, \ldots, k$. Each entry of Θ_i is a real value between 0 and 1. Note that there is no condition for Θ_i to be stochastic; we require only that each entry of Θ_i be in the interval $(0, 1)$ to represent a probability. Then the probability $P[u, v]$ of an edge (u, v) is defined as the product of link affinities corresponding to the values of individual attributes:

$$P[u, v] = \prod_{i=1}^{k} \Theta_i \left[a_i(u), a_i(v)\right], \tag{2.1}$$

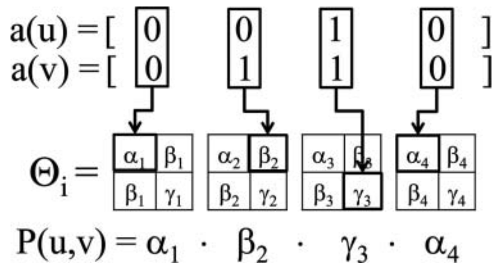


Figure 2. Schematic representation of the multiplicative attribute graphs (MAG) model. Given a pair of nodes u and v with the corresponding binary attribute vectors $a(u)$ and $a(v)$, the probability of an edge $P[u, v]$ is the product over the entries of attribute link-affinity matrices Θ_i, where values of $a_i(u)$ and $a_i(v)$ "select" appropriate entries (row/column) of Θ_i. Note that this particular model represents an undirected graph by making each link-affinity matrix Θ_i symmetric. However, the MAG model generally represents directed graphs.

where $a_i(u)$ denotes the value of the ith attribute of node u. Note that edges appear independently with probability determined by the values of node attributes and link-affinity matrices Θ_i.

Thus, the MAG model M is fully specified by a tuple $M(V, \{a(u)\}, \{\Theta_i\})$, where V is a set of vertices, $\{a(u)\}$ (for each $u \in V$) is a set of vectors capturing attribute values of node u, and $\{\Theta_i\}$ (for $i = 1, \ldots, k$) is a set of link-affinity matrices. Figure 2 illustrates the model.

One can think of the MAG model in the following sense. In order to construct a social network, we ask each node u a series of multiple-choice questions, and the attribute vector $a(u)$ stores the answers of node u to these questions. Then the answers of nodes u and v to the question i determine an entry of the link-affinity matrix Θ_i. In other words, u's answer selects a row, and v's answer selects a column. Assuming that the questions are chosen so that answers are uncorrelated, the product over the entries of the link-affinity matrices Θ_i gives the probability of the edge between u and v.

The choice of multiplicatively combining entries of Θ_i is very natural. In particular, the social network literature defines the concept of Blau space [McPherson 83, McPherson and Ranger-Moore 91], whereby socio-demographic attributes act as dimensions. The organizing force in a Blau space is homophily, in that it has been argued that the flow of information between a pair of nodes decreases with the "distance" in the corresponding Blau space. In this way, small pockets of nodes appear and lead to the development of social niches for human activity and social organization. In this respect, multiplication is a natural way to combine node-attribute data (that is, the dimensions of the Blau space) so that even a single attribute can have profound impact on the linking structure (that is, it creates a narrow social niche community).

As we show next, the proposed MAG model is analytically tractable in the sense that we can formally analyze the properties of the model. Moreover, the MAG model is also statistically interesting, since it can account for the heterogeneities in the node population and can be used to study the interaction between properties of nodes and their linking behavior. Moreover, one can pose many interesting statistical inference questions: Given attribute vectors of all nodes and the network structure, how can we estimate the values of link-affinity matrices Θ_i? How can we infer the attributes of unobserved nodes? Or given a network, how can we estimate both the node attributes and the link-affinity matrices Θ_i? However, the focus of the present paper is the mathematical analysis of the model. For readers interested in the MAG model parameter inference, we point to our follow-up work, where we have developed a method for determining the attribute vectors $a(u)$ and the link-affinity matrices Θ_i for a given network [Kim and Leskovec 11a].

2.3. Simplified Version of the Model

Next we describe a simplified version of the model that we then mathematically analyze in the further sections of the paper. First, while the general MAG model applies to directed networks, we consider the undirected version of the model by requiring each Θ_i to be symmetric. Second, we assume binary attributes, and thus link-affinity matrices Θ_i have two rows and two columns. Third, to further reduce the number of parameters, we also assume that the link-affinity matrices for all attributes are the same, so $\Theta_i = \Theta$ for all i. Putting the three conditions together, we get that $\Theta = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix}$. In other words,

$$\Theta[0,0] = \alpha, \quad \Theta[0,1] = \Theta[1,0] = \beta, \quad \Theta[1,1] = \gamma, \quad \text{for } 0 \le \alpha, \beta, \gamma \le 1.$$

Furthermore, all our results will hold for $\alpha > \beta > \gamma$, which essentially corresponds to the core–periphery structure of the network. However, we note that the assumption $\alpha > \beta > \gamma$ is natural, since most large real-world networks have a common "onion"-like core–periphery structure [Leskovec et al. 08, Leskovec et al. 09, Leskovec et al. 10].

Lastly, we also assume a simple generative model of node-attribute vectors. We consider that each binary attribute vector is generated by k independently and identically distributed (i.i.d.) coin flips with bias μ. That is, we use an i.i.d. Bernoulli distribution parameterized by μ to model attribute vectors such that the probability that the ith attribute of node u takes the value 0 is $P\left(a_i(u) = 0\right) = \mu$ for $i = 1, \ldots, k$ and $0 < \mu < 1$.

Putting it all together, the simplified MAG model $M(n, k, \mu, \Theta)$ is fully specified by six parameters: n is the number of nodes, k is the number of attributes of each node, μ is the probability that an attribute takes the value 0, and $\Theta = [\alpha\ \beta; \beta\ \gamma]$ (where $\alpha > \beta > \gamma$) specifies the attribute link-affinity matrix.

We now study the properties of the random graphs that result from the MAG model $M(n, k, \mu, \Theta)$, where every unordered pair of nodes (u, v) is independently connected with probability $P[u, v]$ defined in (2.1). Since the probability of an edge exponentially decreases in k, the most interesting case occurs when $k = \rho \log_2 n$ for some constant ρ. The choice of $k = \rho \log_2 n$ is motivated by the fact that the effective number of dimensions required to represent networks is of order $\log n$ [Bonato et al. 10]. For convenience, throughout the paper, we will denote $\log_2(\cdot)$ by $\log(\cdot)$, while writing $\log_e(\cdot)$ as $\ln(\cdot)$.

2.4. Connections to Other Models of Networks

We note that our MAG model belongs to a general class of *latent space network models*, where nodes have some discrete or continuous latent attributes and the probability of a pair of nodes forming a link depends on the values of the latent attribute of the two nodes. For example, the latent space model [Hoff and Raftery 02] assumes that nodes reside in a d-dimensional Euclidean space and the probability of an edge between the pair of nodes depends on the Euclidean distance between the latent positions of the two nodes. Similarly, in the random dot product graphs model [Young and Scheinerman 07], the linking probability depends on the inner product between the vectors associated with node positions.

Furthermore, the MAG model generalizes two recent models of network structure. First, the recently developed multifractal network generator [Palla et al. 10] can be viewed as a special case of the MAG model for which the node attribute value distributions as well as the link-affinity matrices are all equal for all attributes.

Moreover, the MAG model also generalizes the Kronecker graphs model [Leskovec et al. 10] in a very subtle way. The Kronecker graphs model takes a small (usually 2×2) initiator matrix K and tensor-powers it k times to obtain a matrix G of size $2^k \times 2^k$, interpreted as the stochastic graph adjacency matrix. One can think of a Kronecker graph model as a special case of the MAG model.

Proposition 2.1. *A Kronecker graph G on 2^k nodes with a 2×2 initiator matrix K is equivalent to the following MAG graph M: Let us number the nodes of M as $0, \ldots, 2^k - 1$. Let the binary attribute vector of a node u of M be a binary representation of its node identifier, and let $\Theta_i = K$. Then individual edge probabilities (u, v) of nodes in G match those in M: $P_G[u, v] = P_M[u, v]$.*

The above observation is interesting for several reasons. First, all results obtained for Kronecker graphs naturally apply to a subclass of MAG graphs where the node's attribute values are the binary representation of its identifier. This means that in a Kronecker graph version of the MAG model, each node has a unique combination of attribute values (that is, each node has a different node identifier), and all attribute value combinations are occupied (that is, node identifiers are in the range $0, \ldots, 2^k - 1$).

Second, the Kronecker graph model can generate only networks whose number of nodes is an integer power of the size of the Kronekcer initiator matrix K ($n = 2^k$). On the other hand, the MAG model provides an important extension in a sense that it does not suffer from this constraint. Our MAG model generates networks with *any* number of nodes.

Third, building on this correspondence between Kronecker and MAG graphs, we also note that the estimates of the Kronecker initiator matrix K nicely transfer to the matrix Θ of the MAG model. For example, the Kronecker initiator matrix $K = [\alpha = 0.98,\ \beta = 0.58,\ \gamma = 0.05]$ accurately models the graph of Internet connectivity, while the global network structure of the Epinions online social network is captured by $K = [\alpha = 0.99,\ \beta = 0.53,\ \gamma = 0.13]$ (see [Leskovec et al. 10]). Thus, in the rest of this paper, we will consider the above values of α, β, and γ as the typical values that the matrix Θ would normally take. In this respect, our assumption of $\alpha > \beta > \gamma$ seems very natural.

Furthermore, the fact that most large real-world networks satisfy $\alpha > \beta > \gamma$ tells us that such networks have recursive core–periphery structure [Leskovec et al. 08, Leskovec et al. 10]. In other words, the network is composed of denser and denser layers of edges as one moves toward the core of the network. Basically, $\alpha > \beta > \gamma$ means that more edges are likely to appear between nodes that share 0's on more attributes, and these nodes form the core of the network. Since more edges appear between pairs of nodes with attribute combination 0–1 than between those with 1–1, there are more edges between the core and the periphery nodes (edges 0–1) than between the nodes of the periphery themselves (edges 1–1).

In the sections that follow we mathematically analyze the properties of the MAG model. We focus mostly on the simplified version. Each section states the main theorem and gives an overview of the proof. We omit the full proofs in the main body of the paper and describe them in the appendix, Section 10.

3. The Number of Edges

In this section, we derive an expression for the expected number of edges in the MAG model. Moreover, this formula can validate not only the assumption

$k = \rho \log n$, but also a substantial social network property, namely the densification power law [Leskovec et al. 05b].

Theorem 3.1. *For a MAG graph $M(n, k, \mu, \Theta)$, the expected number of edges, denoted by m, satisfies*

$$\mathbb{E}[m] = \frac{n(n-1)}{2}\left(\mu^2\alpha + 2\mu(1-\mu)\beta + (1-\mu)^2\gamma\right)^k + n\left(\mu\alpha + (1-\mu)\gamma\right)^k.$$

The expression is divided into two terms. The first term indicates the number of edges between distinct nodes, whereas the second term represents the number of self-edges. If we exclude self-edges, the number of edges will be reduced to the first term.

Before presenting the actual analysis, we define some useful shorthand notation that will be used throughout the paper. First, let V be the set of nodes in the MAG graph $M(n, k, \mu, \Theta)$. We refer to the *weight* of a node u as the number of 0's in its attribute vector, and denote it by $|u|$, that is, $|u| = \sum_{i=1}^{k} \mathbf{1}\{a_i(u) = 0\}$, where $\mathbf{1}\{\cdot\}$ is an indicator function. Additionally, we define W_j to be the set of all nodes with the same weight j, so $W_j = \{u \in V : |u| = j\}$ for $j = 0, 1, \ldots, k$. Similarly, S_j denotes the set of nodes with weight greater than or equal to j, that is, $S_j = \{u \in V : |u| \geq j\}$. By definition, $S_j = \cup_{i=j}^{k} W_i$.

To prove Theorem 3.1, we use the definition of the simplified MAG model and first derive the following two lemmas.

Lemma 3.2. *For distinct $u, v \in V$,*

$$\mathbb{E}[P[u,v] \mid u \in W_i] = \left(\mu\alpha + (1-\mu)\beta\right)^i \left(\mu\beta + (1-\mu)\gamma\right)^{k-i}.$$

Lemma 3.3. *For $u \in V$,*

$$\mathbb{E}[\deg(u) \mid u \in W_i] = (n-1)\left(\mu\alpha + (1-\mu)\beta\right)^i \left(\mu\beta + (1-\mu)\gamma\right)^{k-i} + 2\alpha^i\gamma^{k-i}.$$

Using these lemmas, the outline of the proof of Theorem 3.1 is as follows. Since the number of edges is half the degree sum, all we need to do is to sum $\mathbb{E}[\deg(u)]$ over the degree distribution. However, because $\mathbb{E}[\deg(u)] = \mathbb{E}[\deg(v)]$ if the weights of u and v are the same, we can add up $\mathbb{E}[\deg(u) \mid u \in W_i]$ over the *weight* distribution. Moreover, Theorem 3.1 also points out two important features of the MAG model. First, our earlier assumption that $k = \rho \log n$ for a constant ρ is reasonable on account of the following two corollaries:[1]

Corollary 3.4. *We have $m \in o(n)$ almost surely as $n \to \infty$ if*

$$\frac{k}{\log n} > -\frac{1}{\log\left(\mu^2\alpha + 2\mu(1-\mu)\beta + (1-\mu)^2\gamma\right)}.$$

Corollary 3.5. *We have $m \in \Theta(n^{2-o(1)})$ almost surely as $n \to \infty$ if $k \in o(\log n)$.*

Note that $\log\left(\mu^2\alpha + 2\mu(1-\mu)\beta + (1-\mu)^2\gamma\right) < 0$ because both μ and γ are less than 1. Thus, in order for the MAG model $M(n, k, \mu, \Theta)$ to have a realistic number of edges—for example, more than the number of nodes n—the number of attributes k should be bounded by order $\log n$ from Corollary 3.4. Similarly, since most networks are sparse (that is, $m \ll n^2$), the case of $k \in o(\log n)$ can also be excluded. In consequence, both Corollary 3.4 and Corollary 3.5 provide upper and lower bounds on the number of attributes k. These bounds support our initial assumption of $k = \rho \log n$.

Second, the expected number of edges can be approximately restated as

$$\frac{1}{2} n^{2+\rho\log(\mu^2\alpha+2\mu(1-\mu)\beta+(1-\mu)^2\gamma)},$$

which means that the MAG model obeys the densification power law [Leskovec et al. 05b], one of the properties of networks that grow over time. The densification power law states that $m(t) \propto n(t)^a$ for $a > 1$, where $m(t)$ and $n(t)$ are the numbers of edges and nodes at time t, and a is the densification exponent. For example, an instance of the MAG model with $\rho = 1$, $\mu = 0.5$ (Proposition 2.1) would have the densification exponent $a = \log(|\Theta|)$, where $|\Theta|$ denotes the sum of the entries of Θ.

Proofs of both lemmas and the theorem are fully described in Section 10.1.

4. Connectedness and the Existence of the Giant Component

In the previous section, we observed that the MAG model follows the densification power law, and we gave conditions on the number of edges in the MAG network. In this section, we mathematically investigate another general property

[1] Throughout this paper, "almost surely" or "with high probability" means that some event occurs with probability $1 - o(1)$.

of networks, the existence of a giant connected component. Furthermore, we also examine the situation in which this giant component covers the entire network, which indicates that the network is connected.

We begin with the theorems that the MAG graph has a giant component and further becomes connected.

Theorem 4.1. (Giant component.) *Only one connected component of size $\Theta(n)$ exists in the MAG model $M(n, k, \mu, \Theta)$ almost surely as $n \to \infty$ if and only if*

$$[(\mu\alpha + (1-\mu)\beta)^{\mu} (\mu\beta + (1-\mu)\gamma)^{1-\mu}]^{\rho} \geq \frac{1}{2}.$$

Theorem 4.2. (Connectedness.) *Let the connectedness criterion function of the MAG model $M(n, k, \mu, \Theta)$ be given by*

$$F_{\mathrm{c}}(M) = \begin{cases} (\mu\beta + (1-\mu)\gamma)^{\rho} & \textit{when } (1-\mu)^{\rho} \geq \frac{1}{2}, \\ [(\mu\alpha + (1-\mu)\beta)^{\nu} (\mu\beta + (1-\mu)\gamma)^{1-\nu}]^{\rho} & \textit{otherwise,} \end{cases}$$

where ν is a solution of

$$\left[\left(\frac{\mu}{\nu}\right)^{\nu} \left(\frac{1-\mu}{1-\nu}\right)^{1-\nu}\right]^{\rho} = \frac{1}{2} \quad \textit{in } (0, \mu).$$

Then the MAG model $M(n, k, \mu, \Theta)$ is connected almost surely as $n \to \infty$ if $F_{\mathrm{c}}(M) > \frac{1}{2}$. In contrast, the MAG model $M(n, k, \mu, \Theta)$ is disconnected almost surely as $n \to \infty$ if $F_{\mathrm{c}}(M) < \frac{1}{2}$.

To prove the above theorems, we first define the monotonicity property of the MAG model.

Theorem 4.3. (Monotonicity.) *For $u, v \in V$,*

$$P\left[u, v \mid |u| = i\right] \leq P\left[u, v \mid |u| = j\right]$$

if $i \leq j$.

Theorem 4.3 ultimately demonstrates that a node of larger weight (defined as the number of zeros in its attribute vector) is more likely to be connected with other nodes. In other words, a node of large weight plays a “core” role in the network, whereas a node of small weight is regarded as “peripheral.” This feature of the MAG model has direct effects on the connectedness as well as on the existence of a giant component.

By the monotonicity property, the minimum degree is likely to be the degree of the minimum-weight node. Therefore, the disconnectedness could be proved by showing that the expected degree of the minimum-weight node is too small to be connected with any other node. Conversely, if this lowest degree is large enough, say $\Omega(\log n)$, then any subset of nodes would be connected with the other part of the graph. Thus, to show the connectedness, the degree of the minimum-weight node must be inspected, using Lemma 3.3.

Note that the criterion in Theorem 4.2 is separated into two cases depending on μ, which tells whether the expected number $\mathbb{E}[|W_0|]$ of weight-0 nodes is greater than 1, because $|W_j|$ is a binomial random variable. If this expectation is greater than 1, then the minimum weight is likely to be close to 0, that is, $O(1)$.

Therefore, the condition for connectedness actually depends on the minimum-weight node. In fact, the proof of Theorem 4.2 is accomplished by computing the expected degree of this minimum-weight node and using some techniques introduced in [Mahdian and Xu 07]. Refer to Section 10.2 for the full proof.

A similar argument also works to explain the existence of a giant component. Instead of focusing on the minimum-weight node, Theorem 4.1 shows that the existence of the $\Theta(n)$ component relies on the degree of the *median*-weight node. We intuitively understand this in the following way. Consider that we delete from the network the nodes of degree smaller than the median degree. If the degree of the median-weight node is large enough, then the remaining half of the network is likely to be connected. The connectedness of this half-network implies the existence of a $\Theta(n)$ component, the size of which is at least $n/2$. In the proof, we actually examine the degrees of nodes of three different weights: μk, $\mu k + k^{1/6}$, and $\mu k + k^{2/3}$. The existence of a $\Theta(n)$ component is determined by the degrees of these nodes.

However, the existence of a $\Theta(n)$ component does not necessarily indicate that it is a unique giant component, since there might be another $\Theta(n)$ component. Therefore, to prove Theorem 4.1 more strictly, the uniqueness of the $\Theta(n)$ component has to follow the existence of it. We can prove the uniqueness by showing that if there are two connected subgraphs of size $\Theta(n)$, then they are connected to each other almost surely.

The complete proofs of these three theorems can be found in Section 10.2.

5. Diameter of the MAG Network

The diameter of real-world networks is usually small, even though the number of nodes can grow large. More interestingly, as the network grows, the diameter tends to shrink [Leskovec et al. 05b]. We can show that networks arising from the

MAG model also exhibit this property. We apply similar ideas as in [Mahdian and Xu 07].

Theorem 5.1. *If* $(\mu\beta + (1-\mu)\gamma)^\rho > \frac{1}{2}$*, then the MAG model* $M(n, k, \mu, \Theta)$ *has a constant diameter almost surely as* $n \to \infty$*.*

This theorem does not specify the exact diameter, but under the given condition, it guarantees a bounded diameter even though $n \to \infty$ using the following lemmas (recall that we defined $S_{\lambda k}$ as $\{u \in V : |u| \geq \lambda k\}$ in Section 3).

Lemma 5.2. *If*

$$(\mu\beta + (1-\mu)\gamma)^\rho > \frac{1}{2} \quad \textit{for } \lambda = \frac{\mu\beta}{\mu\beta + (1-\mu)\gamma},$$

then $S_{\lambda k}$ *has a constant diameter almost surely as* $n \to \infty$*.*

Lemma 5.3. *If*

$$(\mu\beta + (1-\mu)\gamma)^\rho > \frac{1}{2} \quad \textit{for } \lambda = \frac{\mu\beta}{\mu\beta + (1-\mu)\gamma},$$

then all nodes in $V \backslash S_{\lambda k}$ *are directly connected to* $S_{\lambda k}$ *almost surely as* $n \to \infty$*.*

By Lemma 5.3, we can conclude that the diameter of the entire graph is limited to ($2+$ diameter of $S_{\lambda k}$). Since by Lemma 5.2, the diameter of $S_{\lambda k}$ is constant almost surely under the given condition, the actual diameter is also constant.

The proofs are presented in Section 10.3.

6. Log-Normal Degree Distribution

In this section, we analyze the degree distribution of the simplified MAG model. In our analysis we exclude the self-edges not only because computations become simple but also because self-edges have little effect on the degree distribution of the network. Depending on values of Θ, the MAG model produces graphs of various degree distributions. For instance, since the network becomes a sparse Erdős–Rényi random graph if $\alpha \approx \beta \approx \gamma < 1$, the degree distribution will approximately follow the binomial distribution. Similarly, if μ is close to 0 or 1, then the MAG graph again becomes an Erdős–Rényi random graph with edge probability $p = \alpha$ (when $\mu \approx 1$) or γ (when $\mu \approx 0$). Another extreme example is the case that $\alpha \approx 1$ and $\mu \approx 1$. Then the network will be close to a complete graph, which represents a degree distribution different from a sparse Erdős–Rényi random graph.

For these reasons, we will focus on particular ranges of values of parameters μ and Θ. For μ we assume that its value is bounded away from 0 and 1. With regard to Θ, we assume that a reasonable configuration space for Θ is one for which

$$\frac{\mu\alpha + (1-\mu)\beta}{\mu\beta + (1-\mu)\gamma}$$

is between 1.6 and 3. Our condition on Θ can be supported by real examples in [Leskovec and Faloutsos 07]. For example, the particular Kronecker graph presented in Section 2 has the value of this ratio equal to 2.44. Also note that the condition on Θ is crucial for us, since in the analysis we use that

$$\left(\frac{\mu\alpha + (1-\mu)\beta}{\mu\beta + (1-\mu)\gamma}\right)^x$$

grows faster than a polynomial function of x. If

$$\frac{\mu\alpha + (1-\mu)\beta}{\mu\beta + (1-\mu)\gamma}$$

is close to 1, we cannot make use of this fact.

Assuming these conditions on μ and Θ, we obtain the following theorem about the degree distribution of the MAG model.

Theorem 6.1. *In the MAG model $M(n, k, \mu, \Theta)$ that follows the above assumptions, if*

$$[(\mu\alpha + (1-\mu)\beta)^\mu \, (\mu\beta + (1-\mu)\gamma)^{1-\mu}]^\rho > \frac{1}{2},$$

then the tail p_d of the degree distribution follows a log-normal distribution, specifically,

$$\ln \mathcal{N}\left(\ln\left(n(\mu\beta + (1-\mu)\gamma)^k\right) + k\mu \ln R + \frac{k\mu(1-\mu)(\ln R)^2}{2}, \quad k\mu(1-\mu)(\ln R)^2\right),$$

for

$$R = \frac{\mu\alpha + (1-\mu)\beta}{\mu\beta + (1-\mu)\gamma} \quad \textit{as } n \to \infty.$$

In other words, the degree distribution of the MAG model approximately follows a quadratic relationship when plotted on a log-log scale. This result is nice, since some social networks tend to follow the log-normal distribution. For instance, the degree distribution of the *LiveJournal* online social network

[Liben-Nowell et al. 05] as well as the degree distribution of the communication network between 240 million users of Microsoft Instant Messenger [Leskovec and Horvitz 08] tend to follow the log-normal distribution.

To give a brief overview of the proof, we first notice that the expected degree of a node in the MAG model is an exponential function of the node weight by Lemma 3.3. This means that the degree distribution is mainly affected by the distribution of node weights. The node weight follows a binomial distribution, which can be approximated by a normal distribution for sufficiently large k. Because the logarithmic value of the expected degree is linear in the node weight and this weight follows a binomial distribution, the logarithm of the degree approximately follows a normal distribution for large k. This in turn indicates that the degree distribution roughly follows a log-normal distribution.

Note that we required

$$[(\mu\alpha + (1-\mu)\beta)^{\mu} (\mu\beta + (1-\mu)\gamma)^{1-\mu}]^{\rho} > \frac{1}{2},$$

which is related to the existence of a giant component. First, this requirement is perfectly acceptable, because real-world networks have a giant component. Second, as we described in Section 4, this condition ensures that the median degree is large enough. Equivalently, it also indicates that the degrees of half the nodes are large enough. If we refer to the tail of the degree distribution as the degrees of nodes with degrees above the median degree, then we can prove Theorem 6.1. The full proofs for the above analysis are described in Section 10.4.

7. Extension: Power-Law Degree Distribution

So far, we have worked with the simplified version of the MAG model parameterized by only a few variables. Even with these few parameters, the model can generate networks with many properties found in real-world networks. However, regarding the degree distribution, even though the log-normal distribution is one that networks commonly follow, many networks also follow the power-law degree distribution [Faloutsos et al. 99].

In this section, we show that by slightly extending the basic MAG model we can produce networks with the power-law degree distribution. We do not attempt to analyze a general case, but rather we suggest an example of a configuration of the MAG model parameters that leads to networks with power-law degree distributions.

We still maintain the condition that every attribute is binary and independently sampled from a Bernoulli distribution. However, in contrast to the simplified version, we allow each attribute to have a different Bernoulli parameter as well as a different attribute link-affinity matrix associated with it. The formal definition of this model is as follows:

$$P\left(a_i(u)=0\right)=\mu_i, \quad P[u,v]=\prod_{i=1}^{k}\Theta_i\left[a_i(u),a_i(v)\right].$$

The number of parameters here is $4k$, and they consist of μ_i's and Θ_i's for $i=1,2,\ldots,k$. For convenience, we denote this power-law version of the MAG model by $M(n,k,\vec{\mu},\vec{\Theta})$, where $\vec{\mu}=\{\mu_1,\ldots,\mu_k\}$ and $\vec{\Theta}=\{\Theta_1,\ldots,\Theta_k\}$. With these additional parameters, we are able to obtain the power-law degree distribution, as the following theorem describes.

Theorem 7.1. *For $M(n,k,\vec{\mu},\vec{\Theta})$, if*

$$\frac{\mu_i}{1-\mu_i}=\left(\frac{\mu_i\alpha_i+(1-\mu_i)\beta_i}{\mu_i\beta_i+(1-\mu_i)\gamma_i}\right)^{-\delta}$$

for $\delta>0$, then the degree distribution satisfies $p_d \propto d^{-\delta-\frac{1}{2}}$ as $n\to\infty$.

In order to investigate the degree distribution of this model, the following two lemmas are essential.

Lemma 7.2. *The probability that node u in $M(n,k,\vec{\mu},\vec{\Theta})$ has an attribute vector with values $a_i(u)$ (for $i=1,\ldots,k$) is*

$$\prod_{i=1}^{k}(\mu_i)^{\mathbf{1}\{a_i(u)=0\}}(1-\mu_i)^{\mathbf{1}\{a_i(u)=1\}}.$$

Lemma 7.3. *The expected degree of node u in $M(n,k,\vec{\mu},\vec{\Theta})$ is*

$$(n-1)\prod_{i=1}^{k}\left(\mu_i\alpha_i+(1-\mu_i)\beta_i\right)^{\mathbf{1}\{a_i(u)=0\}}\left(\mu_i\beta_i+(1-\mu_i)\,\gamma_i\right)^{\mathbf{1}\{a_i(u)=1\}}.$$

By Lemmas 7.2 and 7.3, if the condition in Theorem 7.1 holds, then the probability that a node has the same attribute vector as node u is proportional to the $(-\delta)$th power of the expected degree of u. In addition, the $(-\frac{1}{2})$th power comes from the Stirling approximation for large k. This roughly explains Theorem 7.1. The complete proof is given in Section 10.5, and the result is also verified by simulation in Figure 5.

8. Simulation Experiments

In the previous sections we performed theoretical analysis of the MAG model. In this section, we use simulation experiments to further demonstrate the properties of networks that arise from the MAG model. First, we generate synthetic MAG graphs with different parameter values to explore how the network properties change as a function of the MAG parameter values. We focus on the change of scalar network properties, such as diameter and the size of the largest connected component of the graph, as a function of the model parameter values. Second, we also run simulations with fixed parameter configurations to check other additional properties of networks under the MAG model that we did not theoretically analyze. In this way, we are able to qualitatively compare networks produced by our model to a given real-world network.

8.1. MAG Model Parameter Space

Here we focus on the simplified version of the MAG model and examine how various network properties vary as a function of parameter settings. We fix all but one parameter and vary the remaining parameters. We vary μ, α, f, and n in the MAG model $M(n, k, \mu, \Theta)$, where α is the first entry of the link-affinity matrix $\Theta = [\alpha\ \beta; \beta\ \gamma]$ and f indicates a scalar factor of Θ, that is, $\Theta = f \cdot \Theta_0$ for a constant $\Theta_0 = [\alpha_0\ \beta_0; \beta_0\ \gamma_0]$.

Figure 3 depicts (a) the number of edges, (b) the fraction of nodes in the largest connected component, and (c) the effective diameter of the network as a function of μ, α, f, and n for a fixed $k = 8$. Here the effective diameter of a network is defined as the 90th percentile of the distribution of shortest path distance between connected pairs of nodes [Leskovec et al. 05b].

First, we notice that the growth of the network in the number of edges is slower than exponential, since the curves on the plot grow sublinearly in Figure 3(a), which has a log-scaled y-axis. Note that based on Theorem 3.1, the network size (the number of edges) is roughly proportional to

$$n^2 \left(\mu^2 \alpha + 2\mu(1-\mu)\beta + (1-\mu)^2\gamma\right)^k .$$

For example, by this formula, the number of edges is proportional to the kth power of f (the eighth power of f in our case). Since the expected number of edges is a polynomial function of each variable (μ, α, f, and n), this sublinear growth on the log scale agrees with our analysis. Furthermore, the larger the degree of the polynomial function for each variable, the closer to the straight line the number-of-edges curve becomes. For instance, the network size grows by a 16th-degree polynomial in μ, whereas it grows as a quadratic function of n. In

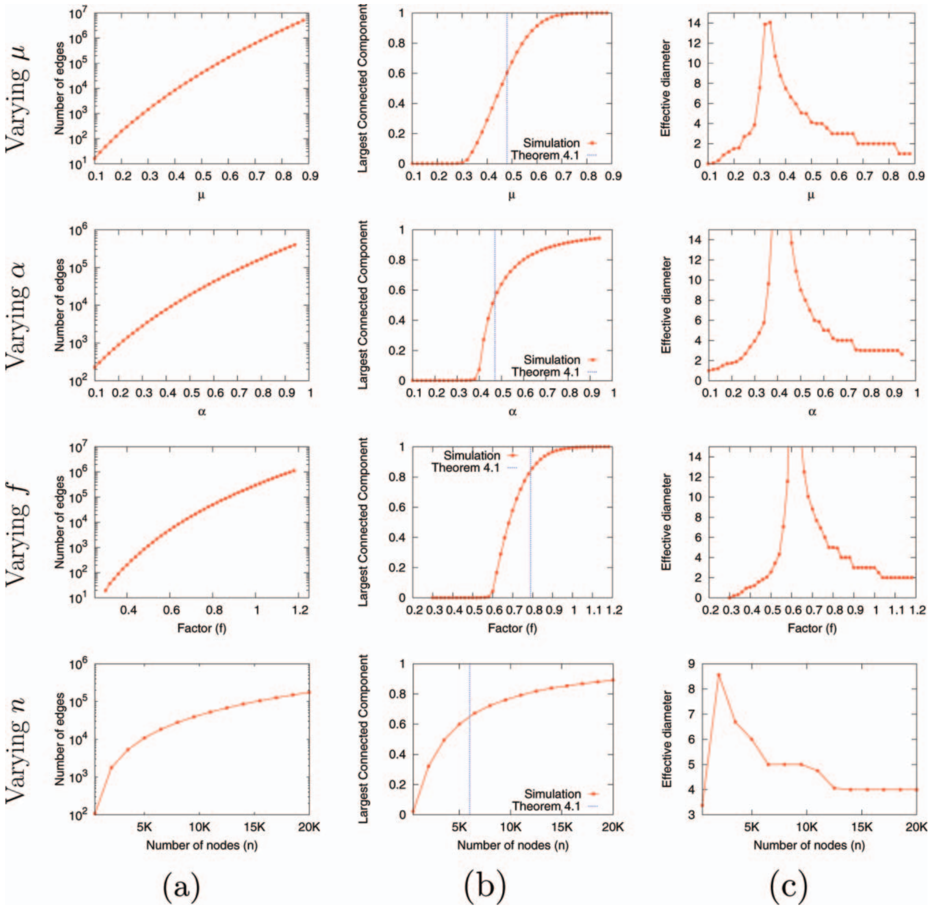


Figure 3. Structural properties of a simplified MAG model $M(n, k, \mu, \Theta)$ when we fix k and vary single parameters one at a time: μ, α, f, or n. As each parameter increases, the synthetic network becomes denser in general, so that a giant connected component emerges and the diameter decreases to approach a constant. (a) Network size, (b) largest connected component, (c) effective diameter (color figure available online).

Figure 3(a), we thus observe that the network size growth over μ is even closer to the exponential curve than that over n.

Second, in Figure 3(b), the size of the largest component shows a sharp thresholding behavior, which indicates a rapid emergence of the giant component. This is very similar to thresholding behaviors observed in other network models such as the Erdős–Rényi random graph and Kronecker models [Erdős and Rényi 60, Leskovec et al. 10]. The vertical line in the middle of each figure

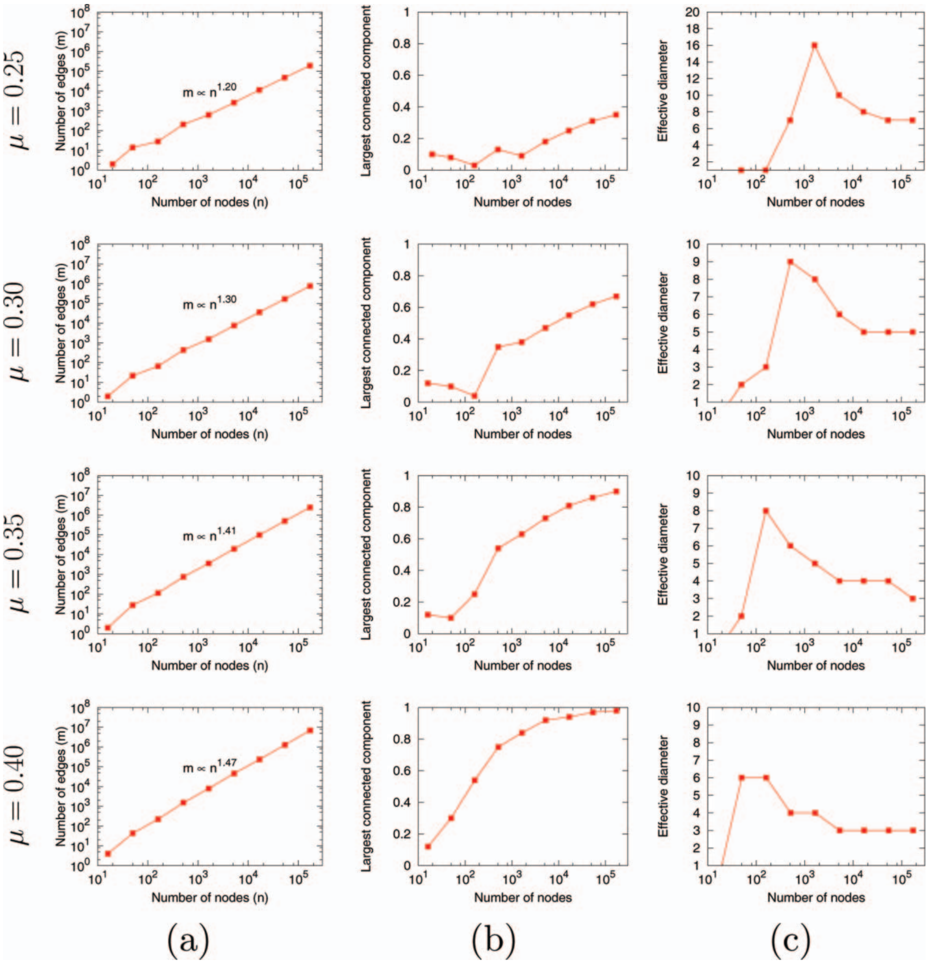


Figure 4. Structural properties of a simplified MAG graph as a function of the number of nodes n for different values of μ (we fix the link-affinity matrix $\Theta = [0.85\ 0.7; 0.7\ 0.15]$ and the ratio $\rho = k/\log n = 0.596$). Observe not only that the relationship between the number of edges and nodes obeys the densification power law, but also that the diameter begins shrinking after the giant component is formed [McGlohon et al. 08]. (a) Network size, (b) largest connected component, (c) effective diameter (color figure available online).

represents the theoretical threshold for the existence of a unique giant connected component (the result of Theorem 4.1). Notice that the theoretical threshold and the simulations agree nearly perfectly. The vertical line intersects the size of the largest connected component close to the point where the largest component contains a bit over 50% of the nodes of the network.

Lastly, while the previous two network properties monotonically change, in Figure 3(c) the effective diameter of the network increases quickly up to about the point where the giant connected component forms and then drops rapidly after that and approaches a constant value. This behavior is in accordance with empirical observations of the "gelling" point where the giant component forms and the diameter starts to decrease in the evolution of real-world networks [Leskovec et al. 05b, McGlohon et al. 08]. Intuitively, this makes sense: When the network is not connected, the effective diameter, which is defined only over connected pairs of nodes, is small. As the network gains more edges, it becomes better connected but looks like a tree, and so the diameter increases. When the network gains even more edges and moves beyond the emergence of the giant connected component, the diameter tends to shrink.

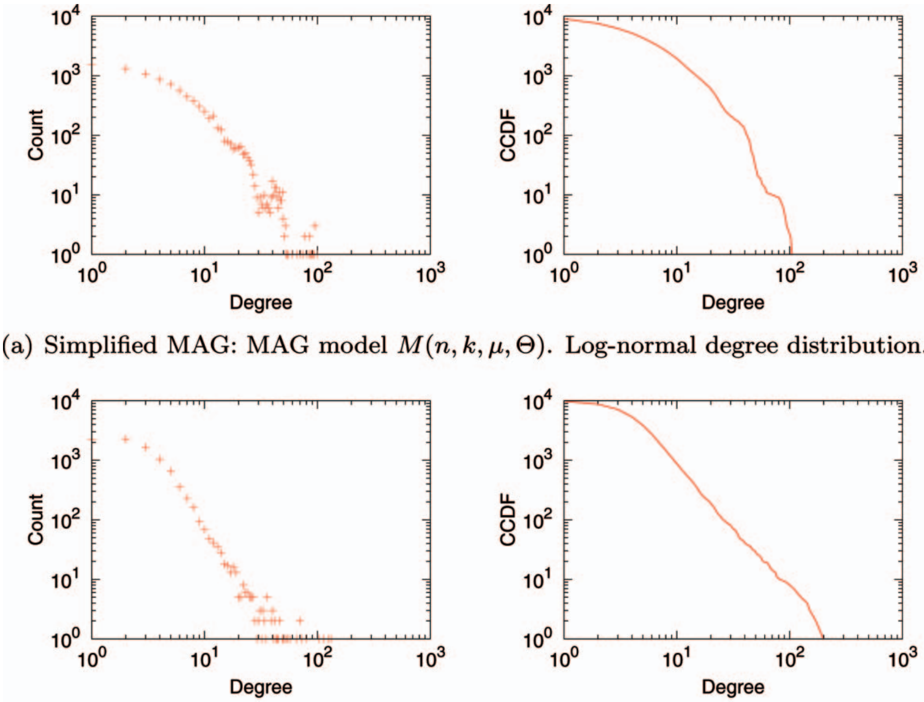
Furthermore, we also performed simulations in which we fix Θ and μ but simultaneously increase both n and k by keeping their ratio constant. Figure 4 plots the change in each network metric (network size, fraction of the largest connected component, and effective diameter) as a function of the number of nodes n for different values of μ. Each plot effectively represents the evolution of the MAG network as the number of nodes grows over time. From the plots, we see that the MAG model follows the densification power law (DPL) and the shrinking diameter property of real-world networks [Leskovec et al. 05b]. Depending on the choice of μ, one can also control for the rate of densification and the diameter.

8.2. Degree Distributions

In addition to the network size, connectivity, and diameter, we also empirically examined the degree distributions of the MAG graph. We already proved that the MAG model can give rise to networks that have either a log-normal or a power-law degree distribution depending on the model parameters. Here we generate the two types of networks and compare their degree distributions.

Figure 5 plots the degree distributions of the two types of MAG model. Figure 5(a) plots the degree distributions of the simplified MAG model $M(n, k, \mu, \Theta)$, while Figure 5(b) plots the degree distribution of the "power-law" version of the MAG model $M(n, k, \vec{\mu}, \vec{\Theta})$ that we introduced in Section 7. For each case, the left plot represents the degree histogram, whereas the right plot shows the *complementary cumulative distribution* (CCDF). While both the histogram and CCDF of a power-law distribution look linear when plotted on a log-log scale, CCDF provides a much better test of whether a distribution follows a power law [Clauset et al. 07].

In Figure 5(a), both raw and CCDF versions of the distribution look parabolic on the log-log scale, which indicates that the MAG model $M(n, k, \mu, \Theta)$ has a



(a) Simplified MAG: MAG model $M(n, k, \mu, \Theta)$. Log-normal degree distribution.

(b) Power-law MAG: $M(n, k, \vec{\mu}, \vec{\Theta})$. Power-law degree distribution.

Figure 5. Degree distributions of simplified and power-law versions of the MAG graph (see Section 7). We plot both the histogram and the CCDF of the degree distribution. The simplified version in figure (a) has a parabolic shape on the log-log scale, which is an indication of a log-normal degree distribution. In contrast, the power-law version in figure (b) shows a straight line on the same scale, which demonstrates a power-law degree distribution (color figure available online).

log-normal degree distribution. On the other hand, in Figure 5(b), both plots exhibit a straight line on the log-log scale, which indicates that the degree distribution of $M(n, k, \vec{\mu}, \vec{\Theta})$ follows a power law. All these experimental results agree with our mathematical analyses in Sections 6 and 7.

8.3. Comparison to Real-world Networks

Also, we qualitatively compare the structural properties of a given real-world network and the corresponding synthetic network generated by the MAG model. In order to compare the real and synthetic networks, we need to determine the appropriate values of the MAG model parameters. This leads to interesting

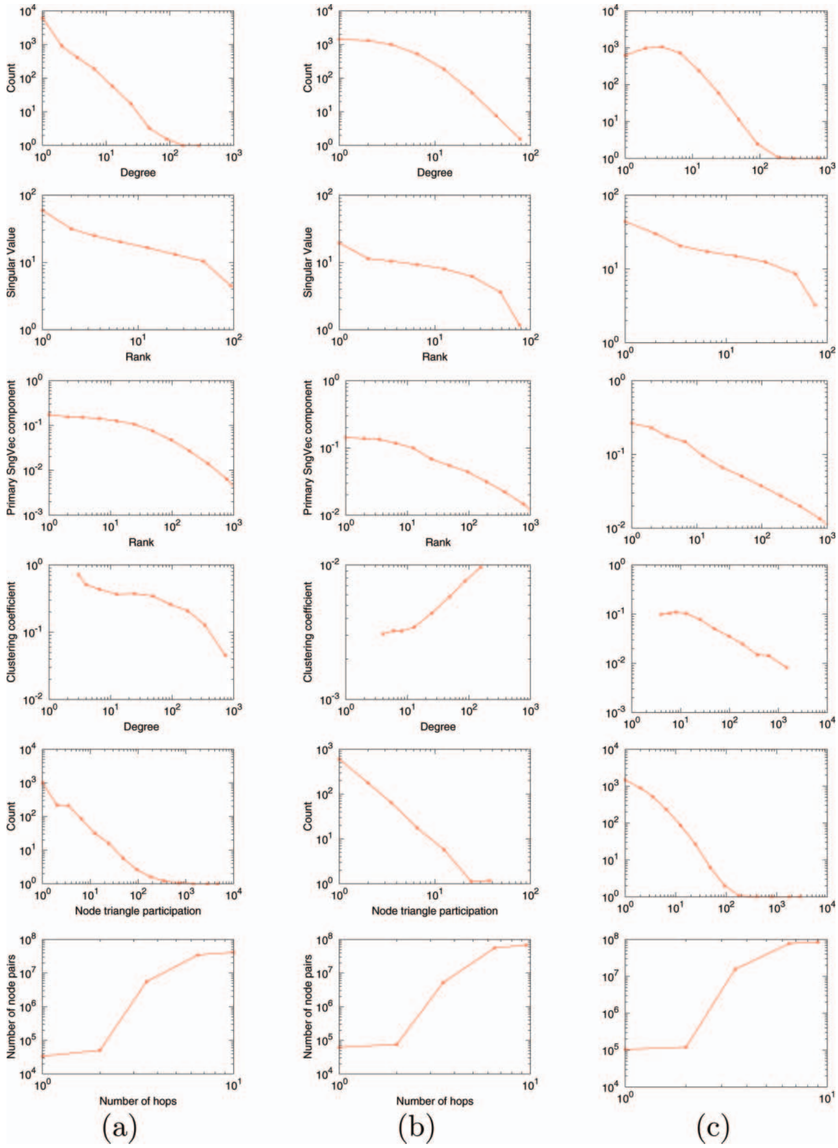


Figure 6. (a) Yahoo!-Flickr network, (b) simplified MAG model, (c) general MAG model. A comparison of network properties between real-world Yahoo!-Flickr online social network, a simplified MAG model network, and a general version of the MAG model. Except for the clustering coefficient, the properties of the MAG model qualitatively resemble those of the Yahoo!-Flickr network even when it is the simplified version in figure (b). Moreover, the general version of the MAG model can represent all six network properties of a shape similar to those of the real-world networks in figure (c) (color figure available online).

questions of how to find optimal MAG model parameters so that the synthetic network resembles the given real-world network. The full resolution of these questions lies beyond the scope of the present paper. In this paper we simply searched by brute force over a relatively small subset of possible MAG parameter settings. For readers interested in how to determine optimal MAG model parameters for a given real-world network, we point to our follow-up work [Kim and Leskovec 11a].

In this paper we manually selected some parameter settings (n, k, μ, Θ) to synthesize the simplified MAG model. We then computed the properties of the MAG model $M(n, k, \mu, \Theta)$ and compared them to the properties of a given real-world network. Our goal is not to claim that these particular parameter values are in any way "optimal" for the given real-world network but rather to demonstrate the flexibility of the MAG model and show that many MAG networks exhibit qualitatively similar properties to those found in real-world networks.

For the real-world network, we use the Yahoo!-Flickr online social network on 10,240 nodes and 44,800 edges. For the simplified MAG model $M(n, k, \mu, \Theta)$, we used $k = 8$, $\mu = 0.45$, $\Theta = [0.85\ 0.30; 0.30\ 0.25]$ with the same number of nodes $n = 10{,}240$. Figures 6(a) and (b) illustrate the following properties of the real-world and corresponding synthetic networks of the simplified MAG model. Each row of the figure plots a separate property:

Row 1: *Degree distribution* is a histogram of the number of edges of a node [Faloutsos et al. 99].

Row 2: *Singular values* indicate the singular values of the adjacency matrix versus their rank [Farkas et al. 01].

Row 3: *Singular vector* represents the distribution of components in the left singular vector associated with the largest singular value [Chakrabarti et al. 04].

Row 4: *Clustering coefficient* represents the degree versus the average (local) clustering coefficient of nodes of a given degree [Watts and Strogatz 98].

Row 5: *Triad participation* indicates the number of triangles to which a node is adjacent. It measures the transitivity in networks [Tsourakakis 08].

Row 6: *Hop plot* shows the number of reachable pairs of nodes as the number of hops. It sketches how quickly the network expands [Palmer et al. 02, Leskovec and Faloutsos 07].

Figure 6 reveals that the plots of properties of the MAG model resemble those of the Yahoo!-Flickr network. Notice the qualitatively similar behavior of

Attribute (i)	μ_i	Θ_i
0	0.60	[0.9999 0.0432; 0.0505 0.9999]
1	0.04	[0.9999 0.9999; 0.9999 0.1506]
2	0.24	[0.9999 0.9999; 0.9999 0.2803]
3	0.17	[0.9999 0.9999; 0.9999 0.2833]
4	0.62	[0.9999 0.0476; 0.0563 0.9999]
5	0.08	[0.9999 0.9999; 0.9999 0.1319]
6	0.57	[0.9999 0.1246; 0.1402 0.9999]
7	0.57	[0.9999 0.1186; 0.1364 0.9999]
8	0.40	[0.9999 0.1757; 0.1535 0.9999]

Table 1. MAG model configuration for Figure 6(c). We let $k = 9$ for this general version.

nearly all properties between Figures 6(a) and (b). The only property for which the simplified MAG model does not match the Yahoo!-Flickr network seems to be the clustering coefficient. Since in real-world networks high-degree nodes tend to have lower clustering, in the simplified MAG model the situation is the reverse: higher-degree nodes also tend to have higher clustering. This is due to the fact that for all attributes we use the same link-affinity matrix Θ, which represents only the core–periphery structure ($\alpha > \beta > \gamma$). Thus, the simplified MAG model can resemble only the overall core–periphery structure of real-world networks [Leskovec et al. 09]. However, networks like the Yahoo!-Flickr network also exhibit local clustering of the edges, where effects of homophily result in formation of tightly knit clusters or communities in networks.

Hence, our hypothesis is that the local clustering of nodes would naturally emerge by mixing core–periphery link-affinity matrices ($\alpha > \beta > \gamma$) and homophily link-affinity matrices ($\alpha, \gamma > \beta$). To investigate this, we also generated the synthetic network with a more general version of the MAG model, $M(n, k, \vec{\mu}, \vec{\Theta})$. Figure 6(c) illustrates the network properties of this general version. We describe the configuration of this version in Table 1. Note that this general version of the model nicely captures the heavy-tailed clustering coefficient distribution that the real-world network shows, which the simplified version cannot do. For the other properties, the general version still exhibits distributions that seem qualitatively similar to those of the real-world network.

By this experiment, we find that the MAG model is capable of representing real-world networks. Furthermore, we verify the flexibility of the MAG model in a sense that it can give rise to networks with different network properties depending on the MAG model parameter configuration.

9. Conclusion

We have presented the multiplicative attribute graph model of real-world networks. Our model considers nodes with categorical attributes as well as the affinity of link formation of each specific attribute. We introduced the attribute link-affinity matrix to represent the affinity of link formation and provide flexibility in the network structure.

The introduced MAG model is both analytically tractable and statistically interesting [Kim and Leskovec 11a]. In this paper, we analytically proved several network properties observed in real-world networks. We proved that the MAG model obeys the densification power law. We also showed both the existence of a unique giant connected component and a small diameter of MAG networks. Furthermore, we showed through mathematical analysis that the MAG model gives rise to networks with either log-normal or power-law degree distribution. Finally, we empirically verified our analytical results by large-scale simulation experiments.

Overall, the MAG model is statistically interesting in a sense that it can represent various types of network structures. Moreover, the MAG model also leads an interesting problem of identifying the structure of a given real-world network in terms of the MAG model parameters. We leave various formulations of the parameter fitting problem for future work. Furthermore, future work includes other kinds of problems such as how to find underlying network structures and missing node attributes where node attributes are partially observed.

10. Appendix

In this appendix we give complete proofs of theorems and lemmas presented earlier in the paper. In particular we prove:

- Section 10.1: The number of edges. We prove Lemmas 3.2, 3.3 and Theorem 3.1.
- Section 10.2: Connectedness and the existence of the giant component. We prove Theorems 4.1, 4.2, and 4.3.
- Section 10.3: Diameter. We prove Lemmas 5.2 and 5.3.
- Section 10.4: Log-normal degree distribution. We prove Theorem 6.1.
- Section 10.5: Power-law degree distribution. We prove Lemmas 7.2, 7.3, and Theorem 7.1.

10.1. The Number of Edges

Here we describe all the proofs with regard to the number of edges in the MAG model $M(n, k, \mu, \Theta)$ as shown in Section 3.

Proof of Lemma 3.2. Let N_{uv}^0 be the number of attributes that take the value 0 in both u and v. For instance, if $a(u) = [0\ 0\ 1\ 0]$ and $a(v) = [0\ 1\ 1\ 0]$, then $N_{uv}^0 = 2$. We similarly define N_{uv}^1 as the number of attributes that take the value 1 in both u and v. Then, $N_{uv}^0, N_{uv}^1 \geq 0$ and $N_{uv}^0 + N_{uv}^1 \leq k$, since k indicates the number of attributes in each node.

By definition of the MAG model, the edge probability between u and v is

$$P[u,v] = \alpha^{N_{uv}^0} \beta^{k-N_{uv}^0-N_{uv}^1} \gamma^{N_{uv}^1}.$$

Since both N_{uv}^0 and N_{uv}^1 are random variables, we need their conditional joint distribution to compute the expectation of the edge probability $P[u,v]$ given the weight of node u. Note that N_{uv}^0 and N_{uv}^1 are independent of each other if the weight of u is given. Let the weight of u be i, which means that $u \in W_i$. Since u and v can share the value 0 only for the attributes where u already takes the value 0, it follows that N_{uv}^0 equivalently represents the number of heads in i coin flips with probability μ. Therefore, N_{uv}^0 follows $\text{Bin}(i, \mu)$. Similarly, N_{uv}^1 follows $\text{Bin}(k-i, 1-\mu)$. Hence their conditional joint probability is

$$P(N_{uv}^0, N_{uv}^1 \mid u \in W_i) = \binom{i}{N_{uv}^0} \mu^{N_{uv}^0} (1-\mu)^{i-N_{uv}^0} \binom{k-i}{N_{uv}^1} \mu^{k-i-N_{uv}^1} (1-\mu)^{N_{uv}^1}.$$

Using this conditional probability, we can compute the expectation of $P[u,v]$ given the weight of u:

$$\begin{aligned}
&\mathbb{E}[P[u,v] \mid u \in W_i] \\
&= \mathbb{E}[\alpha^{N_{uv}^0} \beta^{i-N_{uv}^0} \beta^{k-i-N_{uv}^1} \gamma^{N_{uv}^1} \mid u \in W_i] \\
&= \sum_{N_{uv}^0=0}^{i} \sum_{N_{uv}^1=0}^{k-i} \binom{i}{N_{uv}^0} \binom{k-i}{N_{uv}^1} (\alpha\mu)^{N_{uv}^0} \\
&\qquad \times ((1-\mu)\beta)^{i-N_{uv}^0} (\mu\beta)^{k-i-N_{uv}^1} ((1-\mu)\gamma)^{N_{uv}^1} \\
&= \left[\sum_{N_{uv}^0=0}^{i} \binom{i}{N_{uv}^0} (\alpha\mu)^{N_{uv}^0} ((1-\mu)\beta)^{i-N_{uv}^0} \right] \\
&\quad \times \left[\sum_{N_{uv}^1=0}^{k-i} \binom{k-i}{N_{uv}^1} (\mu\beta)^{k-i-N_{uv}^1} ((1-\mu)\gamma)^{N_{uv}^1} \right] \\
&= (\mu\alpha + (1-\mu)\beta)^i (\mu\beta + (1-\mu)\gamma)^{k-i}.
\end{aligned}$$

□

Proof of Lemma 3.3. By Lemma 3.2 and the linearity of expectation, we sum this conditional probability over all nodes and obtain the expectation of the degree given the weight of node u. $\square$

Proof of Theorem 3.1. We compute the expected number of edges, written as $\mathbb{E}[m]$, by adding up the degrees of all nodes described in Lemma 3.3:

$$\begin{aligned}
\mathbb{E}[m] &= \mathbb{E}[\frac{1}{2}\sum_{u\in V}\deg(u)] \\
&= \frac{1}{2}n\sum_{j=0}^{k}P(W_j)\mathbb{E}[\deg(u)\mid u\in W_j] \\
&= \frac{1}{2}n\sum_{j=0}^{k}\binom{k}{j}\mu^j(1-\mu)^{k-j}\mathbb{E}[\deg(u)\mid u\in W_j] \\
&= \frac{1}{2}n\sum_{j=0}^{k}\binom{k}{j}((n-1)\left(\mu\alpha+(1-\mu)\beta\right)^j\left(\mu\beta+(1-\mu)\gamma\right)^{k-j} \\
&\qquad\qquad + 2\alpha^j\mu^j\gamma^{k-j}(1-\mu)^{k-j}) \\
&= \frac{n(n-1)}{2}\left(\mu^2\alpha+2\mu(1-\mu)\beta+(1-\mu)^2\gamma\right)^k + n\left(\mu\alpha+(1-\mu)\gamma\right)^k .
\end{aligned}$$

$\square$

Proof of Corollary 3.4. Suppose that $k=\left(\epsilon-\frac{1}{\log\zeta}\right)\log n$ for $\zeta=\mu^2\alpha+2\mu(1-\mu)\beta+(1-\mu)^2\gamma$ and $\epsilon>0$. By Theorem 3.1, the expected number of edges is $\Theta\left(n^2\zeta^k\right)$. Note that $\log\zeta<0$, since $\zeta<1$. Therefore, the expected number of edges is

$$\Theta(n^2\zeta^k)=\Theta\left(\zeta^{k+\frac{2\log n}{\log\zeta}}\right)=\Theta(n^{1+\epsilon\log\zeta})=o(n).$$

$\square$

Proof of Corollary 3.5. Under the situation that $k\in o(\log n)$, the expected number of edges is

$$\Theta(n^2\zeta^k)=\Theta(n^{2+(\frac{k}{\log n})\log\zeta})=\Theta(n^{2+o(1)\log\zeta})=\Theta(n^{2-o(1)}).$$

$\square$

10.2. Connectedness and the Existence of the Giant Component

Since Theorem 4.3 is used to prove other theorems, we begin with the proof of it.

Proof of Theorem 4.3. If $j \geq i$, for any $v \in W_i$, we can generate a node $v^{(j)} \in W_j$ from v by flipping $(j-i)$ attribute values that originally take 1 in v. For example, if $a(v) = [0\ 1\ 1\ 0]$, then $a(v^{(3)}) = [0\ 0\ 1\ 0]$ or $[0\ 1\ 0\ 0]$. Hence, $P[u, v^{(j)}] \geq P[u, v]$ for $v \in W_i$.

Here we note that

$$\mathbb{E}[P[u, v^{(j)}] \mid v \in W_i] = \mathbb{E}[P[u, v^{(j)}] \mid v^{(j)} \in W_j],$$

because each $v^{(j)}$ can be generated by $\binom{j}{i}$ different $a(v)$ sets with the same probability. Therefore,

$$\begin{aligned}\mathbb{E}[P[u, v] \mid v \in W_j] &= \mathbb{E}[\mathbb{E}[P[u, v^{(j)}] \mid v \in W_i]] \\ &\geq \mathbb{E}[\mathbb{E}[P[u, v] \mid v \in W_i]] = \mathbb{E}[P[u, v] \mid v \in W_i].\end{aligned}$$

□

Now we introduce the theorem that plays a key role in proving Theorem 4.1 as well as Theorem 4.2.

Theorem 10.1. *Let $|S_j| \in \Theta(n)$ and $\mathbb{E}[P[u, V \backslash u] \mid u \in W_j] \geq c \log n$ as $n \to \infty$ for some j and sufficiently large c. Then, S_j is connected almost surely as $n \to \infty$.*

Proof. Let S' be a subset of S_j such that S' is neither an empty set nor S_j itself. Then the expected number of edges between S' and $S_j \backslash S'$ is

$$\mathbb{E}[P[S', S_j \backslash S'] \mid\ |S'| = k] = k \cdot (|S_j| - k) \cdot \mathbb{E}[P[u, v] \mid u, v \in S_j]$$

for distinct u and v. By Theorem 4.3,

$$\begin{aligned}\mathbb{E}[P[u, v] \mid u, v \in S_j] &\geq \mathbb{E}[P[u, v] \mid u \in S_j, v \in V] \\ &\geq \mathbb{E}[P[u, v] \mid u \in W_j, v \in V \backslash u] \\ &\geq \frac{c \log n}{n}.\end{aligned}$$

Given the size of S' as k, the probability that there exists no edge between S' and $S_j \backslash S'$ is at most $\exp\left(-\frac{1}{2}\mathbb{E}[P[S', S_j \backslash S'] \mid |S'| = k]\right)$ by the Chernoff bound [Ross 05]. Therefore, the probability that S_j is disconnected is bounded as follows:

$$\begin{aligned}
P(S_j \text{ is disconnected}) &\le \sum_{S' \subset S_j, S' \neq \phi, S_j} P(\text{no edge between } S', S_j \backslash S') \\
&\le \sum_{S' \subset S, S' \neq \phi, S_j} \exp\left(-\frac{1}{2}\mathbb{E}[P[S', S_j \backslash S'] \mid |S'|]\right) \\
&\le \sum_{S' \subset S, S' \neq \phi, S_j} \exp\left(-|S'|\left(|S_j| - |S'|\right)\frac{c\log n}{2n}\right) \\
&\le 2 \sum_{1 \le k \le |S_j|/2} \binom{|S_j|}{i} \exp\left(-\frac{c|S_j|\log n}{4n}k\right) \\
&\le 2 \sum_{1 \le k \le |S_j|/2} |S_j|^k \exp\left(-\frac{c|S_j|\log n}{4n}k\right) \\
&\le 2 \sum_{1 \le k \le |S_j|/2} \exp\left(\left(\log|S_j| - \frac{c|S_j|\log n}{4n}\right)k\right) \\
&= 2 \sum_{1 \le k \le |S_j|/2} \exp\left(-k\Theta(\log n)\right) \quad (\text{because } |S_j| \in \Theta(n)) \\
&= 2 \sum_{1 \le k \le |S_j|/2} \left(\frac{1}{n^{\Theta(1)}}\right)^k \\
&\approx \frac{1}{n^{\Theta(1)}} \in o(1)
\end{aligned}$$

as $n \to \infty$. Therefore, S_j is connected almost surely. □

10.2.1. Existence of the Giant Connected Component. Now we turn our attention to the giant connected component. To establish its existence, we investigate $S_{\mu k}$, $S_{\mu k + k^{1/6}}$, and $S_{\mu k + k^{2/3}}$ depending on the situation. The following lemmas tell us the size of each subgraph.

Lemma 10.2. *$|S_{\mu k}| \ge \frac{n}{2} - o(n)$ almost surely as $n \to \infty$.*

Proof. By the central limit theorem [Ross 05], $\frac{|u| - \mu k}{\sqrt{k\mu(1-\mu)}} \sim N(0,1)$ as $n \to \infty$, that is, $k \to \infty$. Therefore, $P(|u| \ge \mu k)$ is at least $\frac{1}{2} - o(1)$, so $|S_{\mu k}| \ge \frac{n}{2} - o(n)$ almost surely as $n \to \infty$. □

Lemma 10.3. *$|S_{\mu k + k^{1/6}}| \in \Theta(n)$ almost surely as $n \to \infty$.*

Proof. By the central limit theorem mentioned in Lemma 10.2,

$$P(\mu k \leq |u| < \mu k + k^{1/6}) \approx \Phi\left(\frac{k^{1/6}}{\sqrt{k\mu(1-\mu)}}\right) - \Phi(0) \in o(1)$$

as $k \to \infty$, where $\Phi(z)$ represents the cumulative distribution function of the standard normal distribution. Since $P(|u| \geq \mu k + k^{1/6})$ is still at least $\frac{1}{2} - o(1)$, the size of $S_{\mu k + k^{1/6}}$ is $\Theta(n)$ almost surely as $k \to \infty$, that is, $n \to \infty$. □

Lemma 10.4. *$|S_{\mu k + k^{2/3}}| \in o(n)$ almost surely as $n \to \infty$.*

Proof. By the Chernoff bound, $P(|u| \geq \mu k + k^{2/3})$ is $o(1)$ as $k \to \infty$. Thus $|S_{\mu k + k^{2/3}}|$ is $o(n)$ almost surely as $n \to \infty$. □

Using the above lemmas, we prove the existence and the uniqueness of the giant connected component under the given condition.

Proof of Theorem 4.1. Existence: First, if

$$[(\mu\alpha + (1-\mu)\beta)^{\mu} (\mu\beta + (1-\mu)\gamma)^{1-\mu}]^{\rho} > \frac{1}{2},$$

then by Lemma 3.3,

$$\begin{aligned}\mathbb{E}[P[u, V\backslash u] \mid u \in W_{\mu k}] &\approx [2[(\mu\alpha + (1-\mu)\beta)^{\mu} (\mu\beta + (1-\mu)\gamma)^{1-\mu}]^{\rho}]^{\log n} \\ &= (1+\epsilon)^{\log n} > c \log n\end{aligned}$$

for some constant $\epsilon > 0$ and $c > 0$. Since $|S_{\mu k}| \in \Theta(n)$ by Lemma 10.2, $S_{\mu k}$ is connected as $n \to \infty$ by Theorem 10.1. In other words, we are able to extract a connected component of size at least $\frac{n}{2} - o(n)$.

Second, when

$$[(\mu\alpha + (1-\mu)\beta)^{\mu} (\mu\beta + (1-\mu)\gamma)^{1-\mu}]^{\rho} = \frac{1}{2},$$

we can apply the same argument for $S_{\mu k + k^{1/6}}$. Because $|S_{\mu k + k^{1/6}}| \in \Theta(n)$ by Lemma 10.3,

$$\begin{aligned}&\mathbb{E}[P[u, V\backslash u] \mid u \in W_{\mu k + k^{1/6}}] \\ &\approx [2[(\mu\alpha + (1-\mu)\beta)^{\mu} (\mu\beta + (1-\mu)\gamma)^{1-\mu}]^{\rho}]^{\log n} \left(\frac{\mu\alpha + (1-\mu)\beta}{\mu\beta + (1-\mu)\gamma}\right)^{(\rho \log n)^{1/6}} \\ &= \left(\frac{\mu\alpha + (1-\mu)\beta}{\mu\beta + (1-\mu)\gamma}\right)^{(\rho \log n)^{1/6}} = (1+\epsilon')^{\rho \log n^{1/6}},\end{aligned}$$

which is also greater than $c \log n$ as $n \to \infty$ for some constant $\epsilon' > 0$. Thus, $S_{\mu k + k^{1/6}}$ is connected almost surely by Theorem 10.1.

Lastly , when

$$[(\mu\alpha + (1-\mu)\beta)^{\mu} (\mu\beta + (1-\mu)\gamma)^{1-\mu}]^{\rho} < \frac{1}{2},$$

for $u \in W_{\mu k + k^{2/3}}$, we have

$$\begin{aligned} &\mathbb{E}[P[u, V\backslash u] \mid u \in W_{\mu k + k^{2/3}}] \\ &\quad \approx [2[(\mu\alpha + (1-\mu)\beta)^{\mu} (\mu\beta + (1-\mu)\gamma)^{1-\mu}]^{\rho}]^{\log n} \\ &\quad \times \left(\frac{\mu\alpha + (1-\mu)\beta}{\mu\beta + (1-\mu)\gamma}\right)^{(\rho \log n)^{2/3}} \\ &\quad = \left[(1-\epsilon'')^{\rho^{-2/3} (\log n)^{1/3}} \left(\frac{\mu\alpha + (1-\mu)\beta}{\mu\beta + (1-\mu)\gamma}\right)\right]^{(\rho \log n)^{2/3}} \end{aligned}$$

is $o(1)$ as $n \to \infty$ for some constant $\epsilon'' > 0$. Therefore, by Theorem 4.3, the expected degree of a node with weight less than $\mu k + k^{2/3}$ is $o(1)$. However, since $S_{\mu k + k^{2/3}}$ is $o(n)$ by Lemma 10.4, $n - o(n)$ nodes have weights less than $\mu k + k^{2/3}$. Hence, most of $n - o(n)$ nodes are isolated, so that the size of the largest component cannot be $\Theta(n)$.

10.2.2. Uniqueness of the Largest Connected Component. We already pointed out that either $S_{\mu k}$ or $S_{\mu k + k^{1/6}}$ is the subset of the $\Theta(n)$ component when the giant connected component exists. Let this component be called H. Without loss of generality, suppose that $S_{\mu k} \subset H$. Then, for any fixed node u,

$$\begin{aligned} P[u, H] &\geq P[u, S_{\mu k}] \qquad (\because S_{\mu k} \subset H) \\ &= |S_{\mu k}| \cdot \mathbb{E}[P[u, v] \mid v \in S_{\mu k}] \\ &\geq |S_{\mu k}| \cdot \mathbb{E}[P[u, v] \mid v \in V\backslash S_{\mu k}] \quad \text{(by Theorem 4.3)} \\ &= \frac{|S_{\mu k}|}{n - |S_{\mu k}|} P[u, V\backslash S_{\mu k}]. \end{aligned}$$

Since $V\backslash H \subset V\backslash S_{\mu k}$, it follows that

$$\mathbb{E}[P[u, V\backslash H]] \leq \mathbb{E}[P[u, V\backslash S_{\mu k}]] \leq \left(\frac{n - |S_{\mu k}|}{|S_{\mu k}|}\right) \mathbb{E}[P[u, H]]$$

holds for every $u \in V$.

Suppose that another connected component H' also contains $\Theta(n)$ nodes. We will deduce a contradiction if H and H' are not connected almost surely as

$n \to \infty$. To determine $\mathbb{E}[P[H, H']]$, we calculate

$$\begin{aligned}\mathbb{E}[P[H, H']] &= |H'| \cdot \mathbb{E}[P[u, H] \mid u \in H'] \\ &\geq \frac{|H'| \cdot |S_{\mu k}|}{n - |S_{\mu k}|} \mathbb{E}[P[u, V \backslash S_{\mu k}] \mid u \in H'] \\ &\geq \frac{|H'| \cdot |S_{\mu k}|}{n - |S_{\mu k}|} \mathbb{E}[P[u, H'] \mid u \in H'] \quad (\text{since } H' \subset V \backslash H \subset V \backslash S_{\mu k}).\end{aligned}$$

However, $\mathbb{E}[P[u, H'] \mid u \in H'] \in \Omega(1)$. Otherwise, since the probability that $u \in H'$ is connected to H' is not greater than $\mathbb{E}[P[u, H'] \mid u \in H']$ by Markov's inequality [Ross 05], u is disconnected from H' almost surely as $n \to \infty$. Therefore, H' includes at least one isolated node almost surely as $n \to \infty$. This is in contradiction to the connectedness of H'.

On the other hand, if $\mathbb{E}[P[u, H'] \mid u \in H'] \in \Omega(1)$, then $\mathbb{E}[P[H, H']] \in \Omega(n)$. In this case, by the Chernoff bound, H and H' are connected almost surely as $n \to \infty$. This is also a contradiction. Therefore, there is no $\Theta(n)$ connected component other than H almost surely as $n \to \infty$. □

10.2.3. Conditions for the Connectedness of a MAG Network. Now we present the proofs for connectedness. Before the main proof, we present and prove an essential lemma.

Lemma 10.5. *If $(1-\mu)^\rho \geq \frac{1}{2}$, then $V_{\min}/k \to 0$ almost surely as $n \to \infty$. Otherwise, if $(1-\mu)^\rho < \frac{1}{2}$, then $V_{\min}/k \to \nu$ almost surely as $n \to \infty$, where ν is a solution of the equation*

$$\left[\left(\frac{\mu}{\nu}\right)^\nu \left(\frac{1-\mu}{1-\nu}\right)^{1-\nu}\right]^\rho = \frac{1}{2}$$

in $(0, \mu)$.

Proof. First, we assume that $(1-\mu)^\rho \geq \frac{1}{2}$, which indicates that $n(1-\mu)^\rho \geq 1$ by definition. Then the probability that $|W_i| = 0$ is at most $\exp(-\frac{1}{2}\mathbb{E}[|W_i|])$ by the Chernoff bound. However, for fixed μ,

$$\mathbb{E}[|W_1|] = n\binom{k}{1}\mu^1(1-\mu)^{k-1} \geq \frac{\mu}{1-\mu}k \in O(k).$$

Therefore, by the Chernoff bound, $P(|W_1| = 0) \to 0$ as $k \to \infty$. This implies that $V_{\min}$ is $o(k)$ almost surely as $n \to \infty$.

Second, we look at the case that $(1-\mu)^\rho < \frac{1}{2}$. For any $\epsilon \in (0, \mu-\nu)$, we use Stirling's approximation,

$$\begin{aligned}
&\mathbb{E}[|W_{(\nu+\epsilon)k}|] \\
&\quad\approx n\binom{k}{(\nu+\epsilon)k}\mu^{(\nu+\epsilon)k}(1-\mu)^{(1-(\nu+\epsilon))k} \\
&\quad\approx \frac{\sqrt{2\pi k}(\frac{k}{e})^k}{\sqrt{2\pi(\nu+\epsilon)k}(\frac{(\nu+\epsilon)k}{e})^{(\nu+\epsilon)k}\sqrt{2\pi(1-(\nu+\epsilon))}\left(\frac{(1-(\nu+\epsilon))k}{e}\right)^{(1-(\nu+\epsilon))k}} \\
&\qquad\times n\mu^{(\nu+\epsilon)k}(1-\mu)^{(1-(\nu+\epsilon))k} \\
&\quad= \frac{n}{\sqrt{2\pi k(\nu+\epsilon)\,(1-(\nu+\epsilon))}}\left[\left(\frac{\mu}{\nu+\epsilon}\right)^{\nu+\epsilon}\left(\frac{1-\mu}{1-(\nu+\epsilon)}\right)^{1-(\nu+\epsilon)}\right]^k.
\end{aligned}$$

Since

$$\left(\frac{\mu}{x}\right)^x\left(\frac{1-\mu}{1-x}\right)^{1-x}$$

is an increasing function of x over $(0,\mu)$, it follows that

$$\left(\frac{\mu}{\nu+\epsilon}\right)^{\nu+\epsilon}\left(\frac{1-\mu}{1-(\nu+\epsilon)}\right)^{1-(\nu+\epsilon)} = (1+\epsilon')\left(\frac{1}{2}\right)^{1/\rho} = (1+\epsilon')n^{-1/k}$$

for some constant $\epsilon' > 0$. Therefore,

$$\mathbb{E}[|W_{(\nu+\epsilon)k}|] = \frac{(1+\epsilon')^k}{\sqrt{2\pi k(\nu+\epsilon)\,(1-(\nu+\epsilon))}}$$

increases exponentially as k increases. By the Chernoff bound, $|W_{(\nu+\epsilon)k}|$ is not zero almost surely as $k\to\infty$, that is, $n\to\infty$.

In a similar way,

$$\mathbb{E}[|W_{(\nu-\epsilon)k}|] = \frac{(1-\epsilon')^k}{\sqrt{2\pi k(\nu-\epsilon)\,(1-(\nu-\epsilon))}}$$

decreases exponentially as k increases. Since $\mathbb{E}[|W_i|] \geq \mathbb{E}[|W_j|]$ if $\mu k \geq i \geq j$, the expected number of nodes with at most weight $(\nu-\epsilon)l$ is less than $(\nu-\epsilon)k\;\mathbb{E}[|W_{(\nu-\epsilon)k}|]$, and its value goes to zero as $k\to\infty$. Hence, by the Chernoff bound, there exists no node of weight less than $(\nu-\epsilon)k$ almost surely as $n\to\infty$.

To sum up, $V_{\min}/k$ goes to ν almost surely as $k\to\infty$, that is, $n\to\infty$. □

Using the above lemma, we prove the condition that the network is connected.

Proof of Theorem 4.2. Let $V_{\min}/k \to t$ for a constant $t \in [0, \mu)$ as $n \to \infty$. If

$$[(\mu\alpha + (1-\mu)\beta)^t (\mu\beta + (1-\mu)\gamma)^{1-t}]^\rho > \frac{1}{2},$$

then by Lemma 3.3,

$$\begin{aligned}\mathbb{E}[P[u, V\backslash u] \mid u \in W_{V_{\min}}] &\approx \mathbb{E}[P[u, V\backslash u] \mid u \in W_{tk}] \\ &\approx [2[(\mu\alpha + (1-\mu)\beta)^t (\mu\beta + (1-\mu)\gamma)^{1-t}]^\rho]^{\log n} \\ &= (1+\epsilon)^{\log n} \\ &\geq c \log n\end{aligned}$$

for some $\epsilon > 0$ and sufficiently large c. Note that $S_{V_{\min}}$ indicates the entire network by definition of $V_{\min}$. Since $|S_{V_{\min}}|$ is $\Theta(n)$, it follows that $S_{V_{\min}}$ is connected almost surely as $n \to \infty$ by Theorem 10.1. Equivalently, the entire network is also connected almost surely $n \to \infty$.

On the other hand, when

$$(\mu\alpha + (1-\mu)\beta)^{\frac{V_{\min}}{\log n}} (\mu\beta + (1-\mu)\gamma)^{\frac{k-V_{\min}}{\log n}} < \frac{1}{2},$$

the expected degree of a node with weight $|V_{\min}|$ is $o(1)$, because from the above relationship,

$$\mathbb{E}[P[u, V\backslash u] \mid u \in W_{V_{\min}}] \approx (1-\epsilon')^{\log n}$$

for some $\epsilon' > 0$. Thus, in this case, some node in $W_{V_{\min}}$ is isolated almost surely, so the network is disconnected. □

10.3. Diameter

Here we first introduce a theorem that plays a key role in proving the constant diameter in the MAG model, and then we present the proofs of Lemmas 5.2 and 5.3 described in Section 5.

Theorem 10.6. [Bollobás 90, Klee and Larmann 81] *For an Erdős–Rényi random graph $G(n, p)$, if $(pn)^{d-1}/n \to 0$ and $(pn)^d/n \to \infty$ for a fixed integer d, then $G(n, p)$ has diameter d with probability approaching* 1 *as $n \to \infty$.*

Proof of Lemma 5.2. Let A^G and A^H be the probabilistic adjacency matrices of random graphs G and H, respectively. If $A^G_{ij} \geq A^H_{ij}$ for every i, j and H has a constant diameter almost surely, then so does G. This can be understood in the following way. To generate a network with A^G, we first generate edges with A^H and further create edges with $(A^G - A^H)$. However, since the edges created in the first step already result in a constant diameter almost surely, it follows that G has a constant diameter.

Note that $\min_{u,v \in S_{\lambda k}} P[u,v] \geq \beta^{\lambda k} \gamma^{(1-\lambda)k}$. Thus it is sufficient to prove that the Erdős–Rényi random graph $G(|S_{\lambda k}|, \beta^{\lambda k}\gamma^{(1-\lambda)k})$ has a constant diameter almost surely as $n \to \infty$. However,

$$\begin{aligned}
&\mathbb{E}[|W_{\lambda k}|]\beta^{\lambda k}\gamma^{(1-\lambda)k} \\
&\quad = n\binom{k}{\lambda k}\mu^{\lambda k}(1-\mu)^{(1-\lambda)k}\beta^{\lambda k}\gamma^{(1-\lambda)k} \\
&\quad \approx \frac{n}{\sqrt{2\pi k\lambda(1-\lambda)}}\left(\frac{\mu\beta}{\lambda}\right)^{\lambda k}\left(\frac{(1-\mu)\gamma}{1-\lambda}\right)^{(1-\lambda)k} \qquad \text{(By Stirling approximation)} \\
&\quad = \frac{n}{\sqrt{2\pi k\lambda(1-\lambda)}}\left(\mu\beta+(1-\mu)\gamma\right)^{k} \qquad \left(\text{because } \lambda = \frac{\mu\beta}{\mu\beta+(1-\mu)\gamma}\right) \\
&\quad = \frac{1}{\sqrt{2\pi k\lambda(1-\lambda)}}\left(2\left(\mu\beta+(1-\mu)\gamma\right)^{\rho}\right)^{\log n} \\
&\quad = \frac{1}{\sqrt{2\pi k\lambda(1-\lambda)}}(1+\epsilon)^{\log n}
\end{aligned}$$

for some $\epsilon > 0$.

Since this value goes to infinity as $n \to \infty$, so does $\mathbb{E}[W_{\lambda k}]$. Therefore, by the Chernoff bound, $|W_{\lambda k}| \geq c\mathbb{E}[W_{\lambda k}]$ almost surely as $n \to \infty$ for some constant c. Then

$$\begin{aligned}
|S_{\lambda k}|\beta^{\lambda k}\gamma^{(1-\lambda)k} &\geq |W_{\lambda k}|\beta^{\lambda k}\gamma^{(1-\lambda)k} \geq c\mathbb{E}[|W_{\lambda k}|]\beta^{\lambda k}\gamma^{(1-\lambda)k} \\
&\approx \frac{c}{\sqrt{2\pi k\lambda(1-\lambda)}}(1+\epsilon)^{\log n}.
\end{aligned}$$

By Theorem 10.6, an Erdős–Rényi random graph

$$G\left(|S_{\lambda k}|, \frac{c(1+\epsilon)^{\log n}}{|S_{\lambda k}|\sqrt{2\pi k\lambda(1-\lambda)}}\right)$$

has diameter at most $(1 + \ln 2/\epsilon)$ almost surely as $n \to \infty$. Thus, the diameter of $G(|S_{\lambda k}|, \beta^{\lambda k}\gamma^{(1-\lambda)k})$ as well as $S_{\lambda k}$ is also bounded by a constant almost surely as $n \to \infty$. □

Proof of Lemma 5.3. For any $u \in V$,

$$\begin{aligned}
P[u, S_{\lambda k}] &\geq \sum_{j=\lambda k}^{k} n \binom{k}{j} \mu^j (1-\mu)^{k-j} \beta^j \gamma^{k-j} \\
&= \sum_{j=\lambda k}^{k} n \binom{k}{j} \lambda^j (1-\lambda)^{k-j} \left(\frac{\mu\beta}{\lambda}\right)^j \left(\frac{(1-\mu)\gamma}{1-\lambda}\right)^{k-j} \\
&= \sum_{j=\lambda k}^{k} n \binom{k}{j} \lambda^j (1-\lambda)^{k-j} \left(\mu\beta + (1-\mu)\gamma\right)^k \\
&= \left(2\left(\mu\beta + (1-\mu)\gamma\right)^\rho\right)^{\log n} \left(\sum_{j=\lambda k}^{k} \binom{k}{j} \lambda^j (1-\lambda)^{k-j}\right).
\end{aligned}$$

By the central limit theorem, $\sum_{j=\lambda k}^{k} \binom{k}{j} \lambda^j (1-\lambda)^{k-j}$ converges to $\frac{1}{2}$ as $k \to \infty$. Therefore, $P[u, S_{\lambda k}]$ is greater than $c \log n$ for a constant c, and then by the Chernoff bound, u is directly connected to $S_{\lambda k}$ almost surely as $n \to \infty$. □

10.4. Log-Normal Degree Distribution

We begin by introducing a general formula for degree distribution in any random graph model.

Theorem 10.7. [Young and Scheinerman 07] *We have*

$$P\left(\deg(u) = d\right) = \int_{u \in V} \binom{n-1}{d} \left(\mathbb{E}[P[u,v]]\right)^d \left(1 - \mathbb{E}[P[u,v]]\right)^{n-1-d} du.$$

If we plug the MAG model $M(n, k, \mu, \Theta)$ into the above theorem, we obtain the following corollary.

Corollary 10.8. *For $E_j = \left(\mu\alpha + (1-\mu)\beta\right)^j \left(\mu\beta + (1-\mu)\gamma\right)^{k-j}$, the probability of degree d in the MAG model $M(n, k, \mu, \Theta)$ is*

$$p_d = \sum_{j=0}^{k} \binom{k}{j} \mu^j (1-\mu)^{k-j} \binom{n-1}{d} E_j^d \left(1 - E_j\right)^{n-1-d}.$$

Proof. We reformulate the statement of Theorem 10.7 as

$$P\left(\deg(u) = d\right) = \sum_{j=0}^{k} P(u \in W_j) \binom{n-1}{d} \left(\mathbb{E}[P[u,v] \mid u \in W_j]\right)^d \times \left(1 - \mathbb{E}[P[u,v] \mid u \in W_j]\right)^{n-1-d},$$

and by applying Lemma 3.2, we obtain the desired formula. □

Now we prove the main theorem on degree distribution for the MAG model $M(n, k, \mu, \Theta)$ using the above corollary.

Proof of Theorem 6.1. To save space, we begin by establishing some shorthand notation for various quantities:

$$\begin{aligned} x &= \mu\alpha + (1-\mu)\beta, \\ y &= \mu\beta + (1-\mu)\gamma, \\ f_j(d) &= \binom{n-1}{d} \left(x^j y^{k-j}\right)^d \left(1 - x^j y^{k-j}\right)^{n-1-d}, \\ g_j(d) &= \binom{k}{j} \mu^j (1-\mu)^{k-j} f_j(d). \end{aligned}$$

By Corollary 10.8, we can restate p_d as $\sum_{j=0}^{l} g_j(d)$.

If most of these terms turn out to be insignificant under our assumptions, the probability p_d will be approximately proportional to one or a few dominant terms. In this case, what we need to do to find j that maximizes $g_j(d) = \binom{k}{j}\mu^j(1-\mu)^{k-j} f_j(d)$ and obtain its approximate formula.

We begin with an approximation of $f_j(d)$. For large n and d, by Stirling approximation,

$$\begin{aligned} f_j(d) &\approx \frac{\sqrt{2\pi n}(n/e)^n \left(x^j y^{k-j}\right)^d \left(1 - x^j y^{k-j}\right)^{n-d}}{\sqrt{2\pi d}(d/e)^d \sqrt{2\pi(n-d)}\left((n-d)/e\right)^{n-d}} \\ &= \frac{1}{\sqrt{2\pi d\left(1 - \frac{d}{n}\right)}} \left(\frac{n x^j y^{k-j}}{d}\right)^d \left(\frac{1 - x^j y^{k-j}}{1 - d/n}\right)^{n-d}. \end{aligned}$$

However, the expected degree of the maximum-weight node is

$$O(n\left(\mu\alpha + (1-\mu)\beta\right)^k),$$

and so is the expected maximum degree. Therefore, the degree d which we are interested in $o(n)$ almost surely as $n \to \infty$, that is, as $k \to \infty$. Therefore,

$$\left(\frac{1 - x^j y^{k-j}}{1 - d/n}\right)^{n-d} \approx \exp\left(-(n-d)x^j y^{k-j} + (n-d)d/n\right) \approx \exp(-n x^j y^{k-j} + d).$$

For sufficiently large k, we can further simplify $g_j(d)$ by normal approximation of the binomial distribution:

$$\begin{aligned}\ln g_j(d) &= \ln \binom{k}{j}\mu^j(1-\mu)^{k-j} + \ln f_j(d)\\ &\approx -\frac{1}{2}\ln(2\pi k\mu(1-\mu)) - \frac{1}{2k\mu(1-\mu)}(j-\mu k)^2 + \ln f_j(d)\\ &\approx C - \frac{1}{2k\mu(1-\mu)}(j-\mu k)^2 - \frac{1}{2}\ln d - d\ln\frac{d}{nx^j y^{k-j}} + d\left(1-\frac{nx^j y^{k-j}}{d}\right)\end{aligned}$$

for some constant C. When $d = nx^\tau y^{k-\tau}$ for $\tau \geq \mu k$ and $R = x/y$, we have

$$\ln g_j(d) \approx C - \frac{1}{2k\mu(1-\mu)}(j-\mu k)^2 - \frac{1}{2}\ln d + d(j-\tau)\ln R + d\left(1-R^{j-\tau}\right).$$

Using $(j-\mu k)^2 = (j-\tau)^2 + (\tau-\mu k)^2 + 2(j-\tau)(\tau-\mu k)$, we obtain

$$\begin{aligned}\ln g_j(d) \approx C_\tau &- \frac{(j-\tau)^2}{2k\mu(1-\mu)} + (j-\tau)\left(d\ln R - \frac{\tau-\mu k}{k\mu(1-\mu)}\right)\\ &+ d\left(1-R^{j-\tau}\right) - \frac{1}{2}\ln d\end{aligned}$$

for

$$C_\tau = C - \frac{(\tau-\mu k)^2}{2k\mu(1-\mu)}.$$

Now considering $g_j(d)$ as a function of j, not of d, we find j that maximizes $g_j(d)$ for $d = nx^\tau y^{k-\tau}$. However, the median weight is approximately equal to μk by the central limit theorem. If we focus on the greater half of the degrees, we can then let $\tau \geq \mu k$. And in this case, since

$$[(\mu\alpha + (1-\mu)\beta)^\mu (\mu\beta + (1-\mu)\gamma)^{1-\mu}]^\rho > \frac{1}{2},$$

we have

$$d \geq [(\mu\alpha + (1-\mu)\beta)^\mu (\mu\beta + (1-\mu)\gamma)^{1-\mu}]^\rho \in \Omega(k).$$

If we differentiate $\ln g_j(d)$ with respect to j, we obtain

$$(\ln g_j(d))' \approx -\frac{j-\tau}{k\mu(1-\mu)} + \left(d\ln R - \frac{\tau-\mu k}{k\mu(1-\mu)}\right) - dR^{j-\tau}\ln R = 0.$$

Because $d \in \Omega(k)$ and $j, \tau \in O(k)$, we can conclude that $R^{j-\tau} \approx 1$ as $n \to \infty$; otherwise, $|(\ln g_j(d))'|$ grows as large as $\Omega(d)$. Therefore, when $j \approx \tau$, $g_j(d)$ is maximized.

Furthermore, since

$$\left|\frac{j-\tau}{2k\mu(1-\mu)}\right| \ll d\ln R \quad \text{as } n \to \infty,$$

the first quadratic term

$$\frac{(j-\tau)^2}{2k\mu(1-\mu)} \quad \text{in } \ln g_j(d)$$

is negligible. As a result, when R is practical (approximately in the range 1.6 to 3.0), $\ln g_{\tau+\Delta}$ will be at most $(\Theta(-d|\Delta|) - \ln g_\tau)$ for $\Delta \geq 1$. After all, g_τ effectively dominates the probability p_d, and so $\ln p_d$ is roughly proportional to $\ln g_\tau$. By assigning

$$\tau = \frac{\ln d - \ln ny^k}{\ln R},$$

we obtain

$$\begin{aligned} \ln p_d &\approx C - \frac{1}{2k\mu(1-\mu)}\left(\frac{\ln d - \ln ny^k}{\ln R} - \mu k\right)^2 - \frac{1}{2}\ln d \\ &= C' - \frac{1}{2k\mu(1-\mu)(\ln R)^2}\left(\ln d - \ln ny^k - k\mu \ln R - \frac{1}{2}k\mu(1-\mu)(\ln R)^2\right)^2 \end{aligned}$$

for some constant C'. Therefore, the degree distribution p_d approximately follows the log-normal distribution, as described in Theorem 6.1. □

10.5. Power-law Degree Distribution

We begin by proving Lemmas 7.2 and 7.3.

Proof of Lemma 7.2. Lemma 7.2 holds because the a_i's are independently distributed Bernoulli random variables. □

Proof of Lemma 7.3. Let us define $P_j(u, v)$ as the edge probability between u and v when we limit consideration only up to the jth attribute:

$$P_j(u,v) = \prod_{i=1}^{j} \Theta_i \left[a_i(u), a_i(v)\right].$$

Thus, what we aim to show is that for a node v,

$$\mathbb{E}[P_k(u,v)] = \prod_{i=1}^{k} \left(\mu_i\alpha_i + (1-\mu_i)\beta_i\right)^{\mathbf{1}\{a_i(u)=0\}} \left(\mu_i\beta_i + (1-\mu_i)\gamma_i\right)^{\mathbf{1}\{a_i(u)=1\}}.$$

When $k = 1$, this is trivially true by Lemma 3.2. When $k > 1$, suppose that the above formula holds for $k = 1, 2, \ldots, k'$. Since

$$P_{k'+1}(u,v) = P_{k'}(u,v)\Theta_{k'+1}\left[a_{k'+1}(u), a_{k'+1}(v)\right],$$

it follows that

$$\begin{aligned}
&\mathbb{E}[P_{k'+1}(u,v)] \\
&\quad = \mathbb{E}[P_{k'}(u,v)]\mathbb{E}[\Theta_{k'+1}[a_{k'+1}(u), a_{k'+1}(v)]] \\
&\quad = \mathbb{E}[P_{k'}(u,v)](\mu_{k'+1}\alpha_{k'+1} + (1-\mu_{k'+1})\beta_{k'+1})^{\mathbf{1}\{a_{k'+1}(u)=0\}} \\
&\qquad \times (\mu_{k'+1}\beta_{k'+1} + (1-\mu_{k'+1})\gamma_{k'+1})^{\mathbf{1}\left\{a_{k'+1}(u)=1\right\}} \\
&\quad = \prod_{i=1}^{k'+1} (\mu_i\alpha_i + (1-\mu_i)\beta_i)^{\mathbf{1}\{a_i(u)=0\}} (\mu_i\beta_i + (1-\mu_i)\gamma_i)^{\mathbf{1}\{a_i(u)=1\}}.
\end{aligned}$$

Therefore, the expected degree formula described in Lemma 7.3 holds for every $k \geq 1$. □

Applying the above lemmas to Theorem 10.7, we prove the degree distribution of $M(n, k, \vec{\mu}, \vec{\Theta})$ as described in Theorem 7.1.

Proof of Theorem 7.1. Before the main argument, we need to define the ordered probability mass of attribute vectors as $p_{(j)}$ for $j = 1, 2, \ldots, 2^k$. For example, if the probability of each attribute vector $(00, 01, 10, 11)$ is respectively 0.2, 0.3, 0.4, and 0.1 when $k = 2$, the ordered probability masses are $p_{(1)} = 0.1$, $p_{(2)} = 0.2$, $p_{(3)} = 0.3$, and $p_{(4)} = 0.4$.

Then by Theorem 10.7, we can express the probability of degree d, written as p_d, as follows:

$$p_d = \binom{n-1}{d} \sum_{j=1}^{2^k} p_{(j)}(E_j)^d(1-E_j)^{n-1-d}, \tag{10.1}$$

where E_j denotes the average edge probability of the node that has the attribute vector corresponding to $p_{(j)}$. If the $p_{(j)}$'s and E_j's are configured so that few terms dominate the probability, we may approximate p_d as

$$\binom{n-1}{d} p_{(\tau)}(E_\tau)^d(1-E_\tau)^{n-1-d}$$

for $\tau = \arg\max_j p_{(j)}\,(E_j)^d\,(1-E_j)^{n-1-d}$. Assuming that this approximation holds, we will propose a sufficient condition for the power-law degree distribution and suggest an example for this condition.

To simplify computations, we propose a condition that $p_{(j)} \propto E_j^{-\delta}$ for a constant δ. Then the jth term is

$$\binom{n-1}{d} p_{(j)}\,(E_j)^d\,(1-E_j)^{n-1-d} \propto ((E_j)^{d-\delta}\,(1-E_j)^{n-1-d}),$$

which is maximized when $E_j \approx \frac{d-\delta}{n-1-\delta}$. Moreover, under this condition, if E_{j+1}/E_j is at least $(1+z)$ for a constant $z > 0$, then

$$\frac{p_{(\tau+\Delta)}\left(E_{\tau+\Delta}\right)^d\left(1-E_{\tau+\Delta}\right)^{n-1-d}}{p_{(\tau)}\left(E_{\tau}\right)^d\left(1-E_{\tau}\right)^{n-1-d}}$$

is $o(1)$ for $\Delta \geq 1$ as $n \to \infty$. Therefore, the τth term dominates equation (10.1).

Next, by the Stirling approximation with the above conditions,

$$\begin{aligned}
p_d &\approx \binom{n-1}{d}\left(\frac{d-\delta}{n-1-\delta}\right)^{d-\delta}\left(\frac{n-1-d}{n-1-\delta}\right)^{n-1-d} \\
&\propto \frac{1}{\sqrt{d(n-1-d)}}\,(d-\delta)^{-\delta}\left(\frac{(n-1)(d-\delta)}{d(n-1-\delta)}\right)^d\left(\frac{n-1}{n-1-\delta}\right)^{n-1-d} \\
&\propto d^{-1/2}\,(d-\delta)^{-\delta}\left(1-\frac{\delta}{d}\right)^d \\
&\approx d^{-\delta-1/2}\exp(-\delta)
\end{aligned}$$

for sufficiently large d and n. Thus, p_d is approximately proportional to $d^{-\frac{1}{2}-\delta}$ for large d as $n \to \infty$.

Finally, we prove that the two conditions for the power-law degree distribution are simultaneously feasible by providing an example configuration. If all the $p_{(j)}$ are distinct and

$$\frac{\mu_i}{1-\mu_i} = \left(\frac{\mu_i\alpha_i + (1-\mu_i)\beta_i}{\mu_i\beta_i + (1-\mu_i)\gamma_i}\right)^{-\delta},$$

then we satisfy the condition $p_{(j)} \propto (E_j)^{-\delta}$ by Lemmas 7.2 and 7.3. On the other hand, if we set $\frac{\mu_i}{1-\mu i} = (1+z)^{-2^i\delta}$, then the other condition, $E_{j+1}/E_j \geq (1+z)$, is also satisfied. Since we are free to configure the μ_i's and Θ_i's independently, the sufficient condition for the power-law degree distribution is satisfied. □

Acknowledgments

We thank Daniel McFarland for helpful discussions and comments. Myunghwan Kim was supported by a Kwanjeong Educational Foundation fellowship. This research was supported in part by NSF grants CNS-1010921, IIS-1016909, LLNL grant DE-AC52-07NA27344, the Albert Yu and Mary Bechmann Foundation, IBM, Lightspeed, Microsoft, and Yahoo.

Research was also supported in part by Intelligence Advanced Research Projects Activity (IARPA) via Air Force Research Laboratory (AFRL) contract number FA8650-10-C-7058. The U.S. government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation herein. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, AFRL, or the U.S. government.

A short version of this paper appeared in the *Proceedings of the Seventh Workshop on Algorithms and Models for the Web Graph (WAW '10).*

References

[Aiello et al. 00] W. Aiello, F. Chung, and L. Lu. "A Random Graph Model for Massive Graphs." In *STOC '00*, pp. 171–180, 2000.

[Airoldi et al. 07] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing. "Mixed Membership Stochastic Blockmodels." *JMLR* 9 (2007), 1981–2014.

[Albert et al. 99] R. Albert, H. Jeong, and A.-L. Barabási. "Diameter of the World-Wide Web." *Nature* 401 (1999), 130–131.

[Barabási and Albert 99] A.-L. Barabási and R. Albert. "Emergence of Scaling in Random Networks." *Science* 286 (1999), 509–512.

[Blum et al. 06] A. Blum, H. Chan, and M. Rwebangira. "A Random-Surfer Web-Graph Model." In *ANALCO '06*, 2006.

[Bodine et al. 10] E. Bodine-Baron, B. Hassibi, and A. Wierman. "Distance-Dependent Kronecker Graphs for Modeling Social Networks." *IEEE THEMES*, 2010.

[Bollobás 90] B. Bollobás. "The Diameter of Random Graphs." *IEEE Trans. Inform. Theory* 36:2 (1990), 285–288.

[Bollobás and Riordan 03] B. Bollobás and O. Riordan. "Mathematical Results on Scale-Free Random Graphs." In *Handbook of Graphs and Networks*, edited by S. Bornholdt and H. Schuster, pp. 1–37. Wiley, 2003.

[Bonato et al. 10] A. Bonato, J. Janssen, and P. Pralat. "The Geometric Protean Model for On-Line Social Networks." In *WAW '10*, 2010.

[Borgs et al. 07] C. Borgs, J. Chayes, C. Daskalakis, and S. Roch. "First to Market Is Not Everything: An Analysis of Preferential Attachment with Fitness." In *STOC '07*, pp. 135–144, 2007.

[Broder et al. 00] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. "Graph Structure in the Web: Experiments and Models." In *WWW '00*, 2000.

[Chakrabarti and Faloutsos 06] D. Chakrabarti and C. Faloutsos. "Graph Mining: Laws, Generators, and Algorithms." *ACM Computing Survey* 38:1 (2006), 2.

[Chakrabarti et al. 04] D. Chakrabarti, Y. Zhan, and C. Faloutsos. "R-mat: A Recursive Model for Graph Mining." In *SDM '04*, 2004.

[Clauset et al. 07] A. Clauset, C. R. Shalizi, and M. E. J. Newman. "Power-Law Distributions in Empirical Data." ArXiv:0706.1062, 2007.

[Cooper and Frieze 03] C. Cooper and A. Frieze. "A General Model of Web Graphs." *RSA* 22:3 (2003), 311–335.

[Erdős and Rényi 60] P. Erdős and A. Rényi. "On the Evolution of Random Graphs." *Publication of the Mathematical Institute of the Hungarian Acadamy of Science* 5 (1960), 17–67.

[Faloutsos et al. 99] M. Faloutsos, P. Faloutsos, and C. Faloutsos. "On Power-Law Relationships of the Internet Topology." In *SIGCOMM '99*, pp. 251–262, 1999.

[Farkas et al. 01] I. J. Farkas, I. Derényi, A.-L. Barabási, and T. Vicsek. "Spectra of 'Real-World' Graphs: Beyond the Semicircle Law." *Phys. Rev. E* 64:2 (2001), 026704.

[Flaxman et al. 04] A. D. Flaxman, A. M. Frieze, and J. Vera. "A Geometric Preferential Attachment Model of Networks." In *WAW '04*, pp. 44–55, 2004.

[Fortunato 10] S. Fortunato. "Community Detection in Graphs." *Physics Reports* 486:3-5 (2010), 75–74.

[Girvan and Newman 02] M. Girvan and M. Newman. "Community Structure in Social and Biological Networks." *Proceedings of the National Academy of Sciences of the United States of America* 99:12 (2002), 7821–7826.

[Gleich and Owen 11] D. F. Gleich and A. B. Owen. "Moment Based Estimation of Stochastic Kronecker Graph Parameters." Arxiv, abs/1106.1674, 2011.

[Hoff and Raftery 02] P. Hoff and A. Raftery. "Latent Space Approaches to Social Network Analysis." *Journal of the American Statistical Association*, 2002.

[Holme 05] P. Holme. "Core-Periphery Organization of Complex Networks." *Physical Review E* 72 (2005), 046111.

[Kim and Leskovec 10] M. Kim and J. Leskovec. "Multiplicative Attribute Graph Model of Real-World Networks." In *WAW '10*, 2010.

[Kim and Leskovec 11a] M. Kim and J. Leskovec. "Modeling Social Networks with Node Attributes Using the Multiplicative Attribute Graph Model." In *UAI '11*, 2011.

[Kim and Leskovec 11b] M. Kim and J. Leskovec. "Network Completion Problem: Inferring Missing Nodes and Edges in Networks." In *SDM '11*, 2011.

[Klee and Larmann 81] V. Klee and D. Larmann. "Diameters of Random Graphs." *Canad. J. Math.* 33 (1981), 618–640.

[Kleinberg 00] J. M. Kleinberg. "Navigation in a Small World." *Nature*, 406 (2000), 6798.

[Kumar et al. 00] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. "Stochastic Models for the Web Graph." In *FOCS '00*, p. 57, 2000.

[Lattanzi and Sivakumar 09] S. Lattanzi and D. Sivakumar. "Affiliation Networks." In *STOC '09*, pp. 427–434, 2009.

[Leskovec 09] J. Leskovec. "Networks, Communities and Kronecker Products." In *CNIKM '09: Proceeding of the 1st ACM International Workshop on Complex Networks Meet Information & Knowledge Management*, pp. 1–2, 2009.

[Leskovec and Faloutsos 07] J. Leskovec and C. Faloutsos. "Scalable Modeling of Real Graphs Using Kronecker Multiplication." In *ICML '07*, 2007.

[Leskovec and Horvitz 08] J. Leskovec and E. Horvitz. "Planetary-Scale Views on a Large Instant-Messaging Network." In *WWW '08: Proceedings of the 17th International Conference on World Wide Web*, 2008.

[Leskovec et al. 05a] J. Leskovec, D. Chakrabarti, J. M. Kleinberg, and C. Faloutsos. "Realistic, Mathematically Tractable Graph Generation and Evolution, Using Kronecker Multiplication." In *PKDD '05*, pp. 133–145, 2005.

[Leskovec et al. 05b] J. Leskovec, J. M. Kleinberg, and C. Faloutsos. "Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations." In *KDD '05*, pp. 177–187, 2005.

[Leskovec et al. 08] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney. "Statistical Properties of Community Structure in Large Social and Information Networks." In *WWW '08*, 2008.

[Leskovec et al. 09] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney. "Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters." *Internet Mathematics* 6:1 (2009), 29–123.

[Leskovec et al. 10] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani. "Kronecker Graphs: An Approach to Modeling Networks." *Journal of Machine Learning Research* 11 (2010), 985–1042.

[Liben-Nowell et al. 05] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. "Geographic Routing in Social Networks." *PNAS* 102:33 (2005), 11623–11628.

[Mahdian and Xu 07] M. Mahdian and Y. Xu. "Stochastic Kronecker Graphs." In *WAW '07*, 2007.

[McGlohon et al. 08] M. McGlohon, L. Akoglu, and C. Faloutsos. "Weighted Graphs and Disconnected Components: Patterns and a Generator." In *KDD*, pp. 524–532, 2008.

[McPherson 83] M. McPherson. "An Ecology of Affiliation." *American Sociological Review* 48:4 (1983), 519–532.

[McPherson and Ranger-Moore 91] M. McPherson and J. R. Ranger-Moore. "Evolution on a Dancing Landscape: Organizations and Networks in Dynamic Blau Space." *Social Forces* 70:1 (1991), 19–42.

[McPherson et al. 01] M. McPherson, L. Smith-Lovin, and J. M. Cook. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:1 (2001), 415–444.

[Milgram 67] S. Milgram. "The Small-World Problem." *Psychology Today* 2 (1967), 60–67.

[Mitzenmacher 04] M. Mitzenmacher. "A Brief History of Generative Models for Power Law and Lognormal Distributions." *Internet Mathematics* 1 (2004), 226–251.

[Mitzenmacher 06] M. Mitzenmacher. Editorial: "The Future of Power Law Research." *Internet Mathematics* 2 (2006), 525–534.

[Palla et al. 10] G. Palla, L. Lovász, and T. Vicsek. "Multifractal Network Generator." *PNAS* 107:17 (2010), 7640–7645.

[Palmer et al. 02] C. R. Palmer, P. B. Gibbons, and C. Faloutsos. "Anf: A Fast and Scalable Tool for Data Mining in Massive Graphs." In *KDD '02*, 2002.

[Rogers and Bhowmik 70] E. M. Rogers and D. K. Bhowmik. "Homophily–Heterophily: Relational Concepts for Communication Research." *Public Opinion Quat.* 34:4 (1970), 523–538.

[Ross 05] S. Ross. *A First Course in Probability*, 7th edition. Upper Saddle River, NJ: Pearson, 2005.

[Tsourakakis 08] C. E. Tsourakakis. "Fast Counting of Triangles in Large Real Networks without Counting: Algorithms and Laws." *ICDM '08*, 2008.

[Wasserman and Pattison 96] S. Wasserman and P. Pattison. "Logit Models and Logistic Regressions for Social Networks." *Psychometrika* 60 (1996), 401–425.

[Watts and Strogatz 98] D. J. Watts and S. H. Strogatz. "Collective Dynamics of 'Small-World' Networks." *Nature* 393 (1998), 440–442.

[Weichsel 62] P. M. Weichsel. "The Kronecker Product of Graphs." *Proceedings of the American Mathematical Society* 13:1 (1962), 37–52.

[Young and Scheinerman 07] S. J. Young and E. R. Scheinerman. *Random Dot Product Graph Models for Social Networks*, Lecture Notes in Computer Science 4863. New York: Springer, 2007.

Myunghwan Kim, Stanford University, Stanford, CA 94305, USA (mykin@stanford.edu)

Jure Leskovec, Stanford University, Stanford, CA 94305, USA (jure@cs.stanford.edu)

Received April 12, 2011; accepted June 3, 2011.