Multiplicative Attribute Graph Model of Real-World Networks*

Myunghwan Kim Jure Leskovec
Stanford University

Abstract

Large scale real-world network data such as social and information networks are ubiquitous. The study of such social and information networks seeks to find patterns and explain their emergence through tractable models. In most networks, and especially in social networks, nodes have a rich set of attributes (*e.g.*, age, gender) associated with them.

Here we present a model that we refer to as the Multiplicative Attribute Graphs (MAG), which naturally captures the interactions between the network structure and the node attributes. We consider a model where each node has a vector of categorical latent attributes associated with it. The probability of an edge between a pair of nodes then depends on the product of individual attribute-attribute affinities. The 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 analyze the degree distribution to show that MAG model can produce networks with either log-normal or power-law degree distributions depending on certain conditions.

1 Introduction

With the emergence of the Web, large online social computing applications have become ubiquitous, which in turn gave rise to a wide range of massive real-world social and information network data such as social networks, computer networks, Internet networks, communication networks, e-mail interactions, Web graphs, and so on. The unifying theme of studying real-world networks is to find patterns of connectivity and explain them through models. The main objective is to answer 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 empirical observations about the structure of networks and the models giving rise to such structures go hand in hand. The empirical analysis of large real-world networks aims to discover common structural properties or patterns, such as heavy-tailed degree distributions [15, 11], local clustering of edges [43, 30], small diameter [3, 28], navigability [36, 22], emergence of community structure [29], 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, the theoretical computer science and physics community have mainly focused on relatively simple "mechanistic" but analytically tractable

*A short version of this paper appeared in *Proceedings of the Seventh Workshop on Algorithms and Models for the Web Graph (WAW'10)* [19].

network models where connectivity patterns observed in the real-world naturally emerge from the model. The prime example in this line of work is the Preferential Attachment model with its many variants [4, 1, 8, 10, 13], which 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 [23], the Small-world model [43, 22], Geometric Random Graphs [17], the Forest Fire model [28], the Random surfer model [5], and models of bipartite affiliation networks [24]. On the other hand, in statistics, machine learning and traditional social network analysis, a different approach to modeling network data has emerged. There the effort is in the development of statistically sound models that consider the structure of the network as well as the features (*e.g.*, age, gender) of nodes and edges in the network. Examples of such models include the Exponential Random Graphs [42], the Stochastic Block Model [2], and the Latent Space Model [18].

"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 a sense that they do not nicely lend themselves to model parameter estimation and are generally too simplistic to model heterogeneities between individual nodes.

On the contrary, "statistical" models are generally analytically intractable and the network properties do not naturally emerge from the model in general. However, these models are usually accompanied by statistical procedures for model parameter estimation and very useful for testing various hypotheses about the interaction of connectivity patterns and the properties 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 [26, 44], which is based on the recursive tensor product of small graph adjacency matrices. The Kronecker graphs model is analytically tractable in a sense that one can analyze global structural properties of networks that emerge from the model [32, 25, 6]. In addition, this model is statistically meaningful because there exists an efficient parameter estimation technique based on maximum likelihood [27, 20]. 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.

Modeling networks with rich node attribute information. Network models investigate edge creation mechanisms, but generally a rich set of attributes is associated with each node. This is especially true in social networks, where not only people's connections but also their characteristics, like age, gender, work place, habits, etc., have been collected. Similarly, various types of profile information is provided by users in online social networks. In this sense, both node characteristics and the network structure need to be considered simultaneously.

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 the nodes or how do we combine node features in an interesting way to obtain probabilities of individual links? While the earlier work on a general class of latent space models formulated such questions, most resulting models were either analytically tractable but statistically uninteresting or statistically very powerful but do not lend themselves to mathematical analysis.

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 the network structure and the node attributes in a clean and tractable manner. We consider a model where each node has a vector of categorical attributes associated with it. Individual attributes of nodes are then combined in order to model the emergence of links. The model allows for rich interaction between node features in a sense that one can simultaneously model features that reflect homophily (*i.e.*, love of the same) as well as heterophily (*i.e.*, love of the different). For example, if people share certain features like hobby, they are more likely to be friends. However, for some other features like gender, people may be more likely to form a relationship with someone with the opposite characteristic. The proposed MAG model is designed to capture both homophily and heterophily that naturally occur in social networks.

We proceed by formulating the model and show that it is both analytically tractable and statistically interesting. In the following sections, we present our mathematical results. Section 3 examines the number of edges and shows that our model naturally obeys the Densification Power Law [28]. Section 4 examines the connectivity of 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. Furthermore, 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 a long-standing debate about how to distinguish the power-law distribution from the log-normal distribution in empirical data [37, 38] and what implications this would make for real-world networks. Also, our results imply that the MAG model 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 the parameters as well as how similar the network looks to real-world networks.

2 Formulating of the Multiplicative Attribute Graph (MAG) model

In this section, we begin with the introduction of the Multiplicative Attribute Graph (MAG) model. Then, we formulate the general version of MAG model and present the simplified version that we analyze throughout this paper. Finally, we investigate the connection to some related works.

2.1 General considerations

We consider a setting where each node u has a vector $a(u)$ of l categorical (*e.g.*, binary) attributes associated with it. For simple examples, one can think of such attribute vectors as a sequence of answers to l yes/no questions such as "Are you female?", "Do you like ice cream?", and so on.

The other 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 account for the homophily of certain features as well as the heterophily of the others by our model. For this mechanism, we associate each attribute i (*i.e.*, i-th question) with an attribute-attribute affinity matrix Θ_i. Each entry of matrix Θ_i represents the affinity depending on the values of the i-th attribute between a pair of nodes. More precisely, $\Theta_i[z_1, z_2]$ indicates the affinity between a pair of nodes, each of which respectively takes value z_1 and z_2 for its i-th attribute. For the binary attribute example in Figure 1, each Θ_i is a 2×2 matrix. To obtain the affinity corresponding to the i-th attribute between node u and v, the values of i-th attribute of both nodes select an appropriate cell of Θ_i.

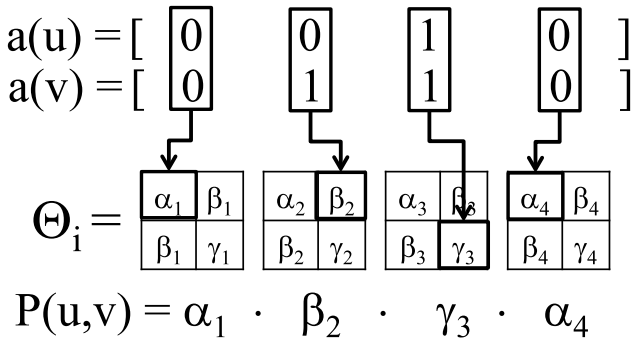
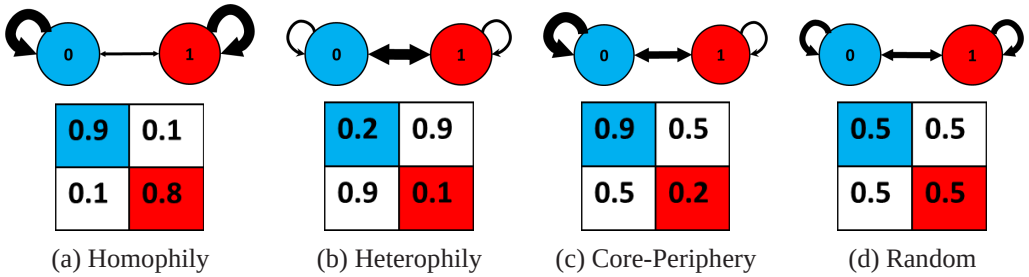


Figure 1: Schematic representation of the Multiplicative Attribute Graphs (MAG) model. Given a pair of nodes u and v with the corresponding binary attribute vectors $a(u)$ and $a(v)$, the probability of edge $P[u, v]$ is the product over the entries of attribute-attribute affinity matrices Θ_i where values of $a_i(u)$ and $a_i(v)$ "select" the appropriate entries (row/column) of Θ_i. Note that this visualized model represents the undirected graph by make each Θ_i symmetric. However, the MAG model in general represents the directed graph.



(a) Homophily (b) Heterophily (c) Core-Periphery (d) Random

Figure 2: Structures in which a node attribute can affect link affinity. The widths of arrows correspond to the affinities towards link formation.

By these affinity matrices, we can capture the various types of structure in real-world social networks. For example, Figure 2 shows four possible linking affinities of a binary attribute. Top figure of each case visualizes the general structure of networks. Each circle represents the group which shares the attribute value and the width of each arrow indicates the affinity of the link formation in the given direction (*e.g.*, the arrow $0 \rightarrow 1$ indicates the affinity of link formation between a node with "0"-value of a given attribute and a node with "1"-value of that attribute.). Then, under each figure, we represent the structure in the form of the affinity matrix.

To investigate one by one, Figure 2(a) shows the homophily (love of the same) attribute affinity and the corresponding affinity matrix Θ. Notice large values on the diagonal entries of Θ, which means that link probability is high when nodes share the same attribute value. 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 value "0" and the other "1". Similarly, Figure 2(b) shows the heterophily (love of the different) affinity, where nodes that do not share the value of the attribute are more likely to link, which gives rise to near-bipartite networks. Also, Figure 2(c) shows the core-periphery affinity, where links are most likely to form between "0" nodes (*i.e.*, members of the core) and least likely to form between "1" nodes (*i.e.*, members of the periphery). Notice that links between the core and the periphery are more likely than the links between the nodes of the

periphery. Additionally, Figure 2(d) illustrates the uniformly random structure that the Erdös-Rényi random graph model generates. By assiging the same value into every entry in each affinity matrix, we can build the MAG model equivalent to the Erdös-Rényi random graphmodel.

From these examples, we notice that the MAG model nicely provides the flexibility in the network structure via the affinity matrices. Although we presented the binary and undirected examples, the MAG model basically allows more complicated structure with larger cardinalities (*e.g.*, 3×3 or 4×4) as well as asymmetric structure through asymmetric affinity matrices.

2.2 The Multiplicative Attributes Graph (MAG) model

Now we formulate a general version of the MAG model. To start with, let each node u have a vector of l categorical attributes and let each attribute have cardinality d_i for $i = 1, 2, \cdots l$. We also have l matrices, $\Theta_i \in d_i \times d_i$ for $i = 1, 2, \cdots l$. Each entry of Θ_i is the affinity of a real value between 0 and 1 [1]. Then, the probability of an edge (u, v), $P[u, v]$, is defined as the multiplication of affinities corresponding to individual attributes, *i.e.*,

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

where $a_i(u)$ denotes the value of i-th attribute of node u. Note that edges appear independently with probability determined by node attributes and matrices Θ_i. Figure 1 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 fo these questions. The answers of nodes u and v on a question i select an entry of matrix Θ_i, *i.e.*, u's answer selects a row and v's answer selects a column. One can thus think of matrices Θ_i's as the attribute-attribute affinity matrices. Assuming that the questions are appropriately chosen so that answers are independent of each other, the product over the entries of matrices Θ_i can be regarded as 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 a concept of Blau-spaces [34, 35] where socio-demographic attributes act as dimensions. Organizing force in Blau space is homophily as 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 (*i.e.*, the dimensions of the Blau space) so that even a single attribute can have profound impact on the linking structure (*i.e.*, it creates a narrow social niche community).

The proposed MAG model model is analytically tractable in a sense that we can formally analyze the properties of the model. Moreover, the MAG model is also statistically interesting as 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 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 matrices Θ_i? However, the focus of this paper is in mathematical analysis and we leave the questions of MAG model parameter estimation for the future work.

[1] Note that there is no condition for Θ_i to be stochastic, we only require each entry of Θ_i to be on interval $(0, 1)$.

2.3 Simplified version of the model

Next we delineate a simplified version of the model that we will 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 node attributes and thus affinity matrices Θ_i have 2 rows and 2 columns. Third, to further reduce the number of parameters, we also assume that the affinity matrices for all attributes are the same, *i.e.*, $\Theta_i = \Theta$ for all i. These three conditions imply that $\Theta = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix}$, *i.e.*, $\Theta[0,0] = \alpha, \Theta[0,1] = \Theta[1,0] = \beta$, and $\Theta[1,1] = \gamma$ for $0 < \alpha, \beta, \gamma < 1$. Furthermore, all our results will hold for $\alpha > \beta > \gamma$. The assumption $\alpha > \beta > \gamma$ is natural since most large real-world networks have a common onion-like "core-periphery" structure [29, 30, 25]. Figure 2(c) exhibits this structure. More precisely, the network is composed from denser and denser layers of edges as one moves towards the core of the network. Basically, $\alpha > \beta > \gamma$ means that more edges are likely to appear between nodes which share value 0 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").

Last, we also assume a simple generative model of node attributes where each binary attribute vector is generated by l independently and identically distributed coin flips with bias μ. That is, we use an *i.i.d.* Bernoulli distribution parameterized by μ to model attribute vectors where the probability that the i-th attribute of a node u takes value 0 is $P(a_i(u) = 0) = \mu$ for $i = 1, \cdots, l$ and $0 < \mu < 1$.

Putting it all together, the MAG model $M(n, l, \mu, \Theta)$ is fully specified by six parameters: n is the number of nodes, l is the number of attributes of each node, μ is the probability that an attribute takes a value of 1, and $\Theta = [\alpha\ \beta; \beta\ \gamma]$ specifies the attribute-attribute affinity matrix.

We now study the properties of the random graphs that result from the $M(n, l, \mu, \Theta)$ where every unordered pair of nodes (u, v) is independently connected with probability $P[u, v]$ defined in Equation (1). Since the probability of an edge exponentially decreases in l, the most interesting case occurs when $l = \rho \log n$ for some constant ρ.[2] This result perfectly agrees that the effective number of dimensions which can represent online social networks is the order of $\log n$ [9].

2.4 Connections to other models

We note that our model belongs to a general class of latent space network models, where nodes have some discrete or continuous valued attributes and the probability of linking depends on the values of attribute of the two nodes. For example, the Latent Space Model [18] assumes that nodes reside in d-dimensional Euclidean space and the probability of an edge depends on the Euclidean distance between the locations of the nodes. Similarly, in Random Dot Product Graphs [45], the linking probability depends on the inner product between the vectors associated with node positions. Furthermore, recently introduced Multifractal Network Generator [39] can also be viewed as a special case of MAG model where the node attribute distribution and the affinity matrix are equal for every attribute.

The MAG model generalizes the Kronecker graphs model [25] in a subtle way. The Kronecker graphs model takes a small (usually 2×2) initiator matrix K and tensor-powers it l times to obtain a matrix G of size $2^l \times 2^l$, interpreted as the stochastic graph adjacency matrix. One can think of a Kronecker graph model as a special case of the MAG model.

[2] Throughout the paper, $\log(\cdot)$ indicates $\log_2(\cdot)$ unless explicitly specified as $\ln(\cdot)$.

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

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 id. This means that in a Kronecker graph version of the MAG model each node has a unique combination of attribute values (*i.e.*, each node has different node id) and all attribute value combinations are occupied (*i.e.*, node ids range $0, \ldots, 2^l - 1$).

Second, building on this correspondence between Kronecker and MAG graphs, we also note that the estimates of the Kronecker initiator matrix K nicely transfer to matrix Θ of MAG model. For example, Kronecker initiator matrix $K = [\alpha = 0.98, \beta = 0.58, \gamma = 0.05]$ accurately models the graph of the internet connectivity, while the global network structure of the Epinions online social network is captured by $K = [\alpha = 0.99, \beta = 0.53, \gamma = 0.13]$ [27]. Thus, in the rest of the paper, we will consider the above values of K as the typical values that the matrix Θ would normally take. In this respect, the assumption of $\alpha > \beta > \gamma$ naturally appears.

In following sections, we analyze the properties of the MAG model. We focus mostly on the simplified version. Each section states the main theorem and gives the overview of the proof. We omit the full proofs in the main body of the paper and describe them in the Appendix.

3 The Number of Edges

In this section, we derive the expression for the expected number of edges in MAG model. Moreover, this formula can valdiate not only the assumption, $l = \rho \log n$, but also a substantial social network property, namely the Densification Power Law.

Theorem 3.1 *For a MAG graph $M(n, l, \mu, \Theta)$, the number of edges, m, satisfies*

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

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

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

To complete the proof of Theorem 3.1, using the definition of the simplified MAG model, we can derive the main lemmas as follows:

Lemma 3.2 *For **distinct** $u, v \in V$, $\mathbb{E}[P[u, v] | u \in W_i] = (\mu\alpha + (1-\mu)\beta)^i (\mu\beta + (1-\mu)\gamma)^{l-i}$.*

Lemma 3.3 *For $u \in V$, $\mathbb{E}[deg(u) | u \in W_i] = (n-1)(\mu\alpha + (1-\mu)\beta)^i (\mu\beta + (1-\mu)\gamma)^{l-i} + 2\alpha^i\gamma^{l-i}$.*

By using these lemmas, the outline of the proof for Theorem 3.1 is as follows. Since the number of edges is half of 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 $\mathbb{E}[deg(u)|u \in W_i]$ over the *weight* distribution, *i.e.*, binomial distribution $Bin(l, \mu)$.

On the other hand, more significantly, Theorem 3.1 can result in two substantial features of MAG model. First, the assumption that $l = \rho \log n$ for a constant ρ can be validated by the next two corollaries.

Corollary 3.3.1 *$m \in o(n)$ with high probability[3] as $n \to \infty$, if $\frac{l}{\log n} > -\frac{1}{\log(\mu^2\alpha + 2\mu(1-\mu)\beta + (1-\mu)^2\gamma)}$.*

Corollary 3.3.2 *$m \in \Theta(n^{2-o(1)})$ with high probability as $n \to \infty$, if $l \in o(\log n)$.*

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 $M(n, l, \mu, \Theta)$ to have a proper number of edges (*e.g.*, more than n), l should be bounded by the order of $\log n$. On the contrary, since most social networks are sparse, $l \in o(\log n)$ case can be also reasonably excluded. In consequence, both Corollary 3.3.1 and Corollary 3.3.2 provide the upper and lower bounds of l for social networks. These bounds eventually support the assumption of $l = \rho \log n$.

Although we do not technically define any process of MAG graph evolution, we can interpret it in the folllowing way. When a new node joins the network, its behavior is governed by the node attribute distribution which is seemingly independent of the network structure. However, in a long term, since the number of attributes grows slowly as the number of nodes increases, the node attributes and the network structure are not independent. This phenomenon is somewhat aligned with the real world. When a new person enters the network, he or she seems to act independently of other people, but people eventually constitue a structured network in the large scale and their behaviors can be categorized into more classes as the network evolves.

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

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

We find that this fact agrees with the Densification Power Law [28], one of the properties of social networks, which indicates $m(t) \propto n(t)^a$ for $a > 1$. For example, an instance of 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 all entries in Θ.

The proofs are fully described in Appendix.

4 Connectivity

In the previous section, we observed that MAG model obeys the Densification Power Law. In this section, we mathematically investigate MAG model for another general property of social networks, the existence of a giant connected component. Furthermore, we also examine the situation where this giant component covers the entire network, *i.e.*, the network is connected.

We begin with the theorems that 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 $M(n, l, \mu, \Theta)$ with high probability as $n \to \infty$, if and only if*

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

[3] It indicates the probability $1 - o(1)$.

Theorem 4.2 (Connectedness) *Let the connectedness criterion function of* $M(n, l, \mu, \Theta)$ *be*

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

where ν *is a solution of* $\left[\left(\frac{\mu}{\nu}\right)^{\nu} \left(\frac{1-\mu}{1-\nu}\right)^{1-\nu}\right]^{\rho} = \frac{1}{2}$ *in* $(0, \mu)$.

Then, $M(n, l, \mu, \Theta)$ *is connected with high probability as* $n \to \infty$*, if* $F_{\mathrm{c}}(M) > \frac{1}{2}$*. In contrast,* $M(n, l, \mu, \Theta)$ *is disconnected with high probability as* $n \to \infty$*, if* $F_{\mathrm{c}}(M) < \frac{1}{2}$.

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

Theorem 4.3 (Monotonicity) *For* $u, v \in V$, $P[u, v||u| = i] \leq P[u, v||u| = j]$ *if* $i \leq j$.

Theorem 4.3 ultimately demonstrates that a node of larger weight 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 the node of small weight is regarded as "periphery". This feature of the MAG model has direct effects on the connectedness as well as on the existence of a giant component.

By the monotonicty 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 should be necessarily inspected, using Lemma 3.3.

Note that the criterion in Theorem 4.2 is separated into two cases depending on μ, which tells whether or not the expected number of weight 0 nodes, $\mathbb{E}[|W_0|]$, is greater than 1, because $|W_j|$ is a binomial random variable. If this expectation is larger than 1, then the minimum weight is likely to be close to 0, *i.e.*, $O(1)$. Otherwise, if $\mathbb{E}[|W_0|] < 1$, the equation of ν describes the ratio of the minimum weight to l as $n \to \infty$. 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 [32].

Similar explanation works for the existence of a giant component. Instead of the minimum weight node, Theorem 4.1 shows that the existence of $\Theta(n)$ component relies on the degree of the *median* weight node. We intuitively understand this in the following way. We might throw away the lower half of nodes by degree. If the degree of the median weight node is large enough, then the half of the network is likely to be connected. The connectedness of this half network implies the existence of $\Theta(n)$ component, the size of which is at least $\frac{n}{2}$. In the proof, we actually examine the degrees of nodes of three different weights: μl, $\mu l + l^{1/6}$, and $\mu l + l^{2/3}$. The existence of $\Theta(n)$ component is determined by the degrees of these nodes.

However, the existence of $\Theta(n)$ 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 $\Theta(n)$ 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 each other with high probability.

The proofs of those three theorems are in Appendix.

5 Diameter

Another property of social networks is that the diameter of the network remains small although the number of nodes grows large. We can show this property in MAG model by applying the similar idea as in [32].

Theorem 5.1 *If* $(\mu\beta + (1-\mu)\gamma)^\rho > \frac{1}{2}$*, then* $M(n, l, \mu, \Theta)$ *has a constant diameter with high probability as* $n \to \infty$*.*

This theorem does not specify the exact diameter, but, under the given condition, it guarantees the bounded diameter even though $n \to \infty$ by using the following lemmas:

Lemma 5.2 *If* $(\mu\beta + (1-\mu)\gamma)^\rho > \frac{1}{2}$*, for* $\lambda = \frac{\mu\beta}{\mu\beta+(1-\mu)\gamma}$*,* $S_{\lambda l}$ *has a constant diameter with high probability as* $n \to \infty$*.*

Lemma 5.3 *If* $(\mu\beta + (1-\mu)\gamma)^\rho > \frac{1}{2}$*, for* $\lambda = \frac{\mu\beta}{\mu\beta+(1-\mu)\gamma}$*, all nodes in* $V \backslash S_{\lambda l}$ *are directly connected to* $S_{\lambda l}$ *with high probability as* $n \to \infty$*.*

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

The proofs are represented in Appendix.

6 Degree Distribution

In this section, we analyze the degree distribution of the simplified MAG model under some reasonable assumptions.[4] Depending on Θ, MAG model produces graphs of various degree distributions. For instance, since the network becomes a sparse Erdös-Rényi random graph if $\alpha \approx \beta \approx \gamma < 1$, the degree distribution will approximately follow the binomial distribution. For another extreme example, in case of $\alpha \approx 1$ and $\mu \approx 1$, the network will be close to a complete graph, which represents a degree distribution different from a sparse Erdös-Rényi random graph. For this reason, we need to narrow down the conditions on μ and Θ as follows. If μ is close to 0 or 1, then the graph becomes an Erdös-Rényi random graph with edge probability $p = \alpha$ (when $\mu \approx 1$) or γ (when $\mu \approx 0$). Since the degree distribution of the Erdös-Rényi random graph is binomial, we will exclude these extreme cases of μ. On the other hand, with regard to Θ, we assume that a reasonable configuration space for Θ would be where $\frac{\mu\alpha+(1-\mu)\beta}{\mu\beta+(1-\mu)\gamma}$ is between 1.6 and 3. For the previous Kronecker graph example, this ratio is actually about 2.44. Our approach for the condition on Θ can be also supported by real examples in [27]. This condition 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 the 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 all these conditions on μ and Θ, we result in the following theorem about the degree distribution.

Theorem 6.1 *In* $M(n, l, \mu, \Theta)$*that follows above assumptions, if*

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

[4] We trivially exclude self-edges not only because computations become simple but also because other models usually do not include them.

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

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

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

In other words, the degree distribution of MAG model approximately follows a quadratic relationship on log-log scale. This result is nice since some social networks follow the log-normal distribution. For instance, the degree distribution of *LiveJournal* network looks more parabolic than linear on log-log scale [31].

In brief, as the expected degree is an exponential function of the node weight by Lemma 3.3, the degree distribution is mainly affected by the distribution of node weights. Since the node weight follows a binomial distribution, it can be approximated to a normal distribution for sufficiently large l. Because the logarithmic value of the expected degree is linear in the node weight and this weight follows a binomial distribution, the log value of degree approximately follows a normal distribution for large l. This eventually indicates that the degree distribution roughly follows a log-normal.

Note that there exists a condition, $\left[(\mu\alpha+(1-\mu)\beta)^{\mu}\left(\mu\beta+(1-\mu)\gamma\right)^{1-\mu}\right]^{\rho}>\frac{1}{2}$, which is related to the existence of a giant component. First, this condition 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 a half of the nodes are large enough. If we refer to the tail of degree distribution as the degrees of nodes with degrees above the median degree, then we can show Theorem 6.1.

The full proofs for this analysis are described in Appendix.

7 Extensions: Power-Law Degree Distribution

So far we have handled the simplified version of MAG model parameterized by only few variables. Even with these few parameters, many well-known properties of social networks can be reproduced. However, regarding to the degree distribution, even though the log-normal is one of the distributions that social networks commonly follow, a lot of social networks also follow the power-law degree distribution [15].

In this section, we show that the MAG model produces networks with the power-law degree distribution by releasing some constraints. We do not attempt to analyze it in a rigorous manner, but give the intuition by suggesting an example of configuration. We still hold the condition that every attribute is binary and independently sampled from 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-attribute affinity matrix associated wit it. The formal definition of this model is as follows:

$$P\left(a_j(u)=0\right)=\mu_j,\ P[u,v]=\prod_{j=1}^{l}\Theta_j\left[a_j(u),a_j(v)\right].$$

The number of parameters here is $4l$, which consist of μ_j's and Θ_j's for $j=1,2,\cdots,l$. For convenience, we denote this power-law version of MAG model as $M(n,l,\vec{\mu},\vec{\Theta})$ where $\vec{\mu}=\{\mu_1,\cdots,\mu_l\}$ and $\vec{\Theta}=\{\Theta_1,\cdots,\Theta_l\}$. With these additional parameters, we are able to obtain the power law degree distribution as the following theorem describes.

Theorem 7.1 *For $M(n,l,\vec{\mu},\vec{\Theta})$, if $\frac{\mu_j}{1-\mu_j}=\left(\frac{\mu_j\alpha_j+(1-\mu_j)\beta_j}{\mu_j\beta_j+(1-\mu_j)\gamma_j}\right)^{-\delta}$ for $\delta>0$, then the degree distribution satisfies $p_k\propto k^{-\delta-\frac{1}{2}}$ 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 a node u in $M(n, l, \vec{\mu}, \vec{\Theta})$ has an attribute vector $a(u)$ is*

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

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

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

By Lemmas 7.2 and 7.3, if the condition in Theorem 7.1 holds, the probability that a node has the same attribute vector as node u is proportional to $(-\delta)$-th power of the expected degree of u. In addition, $(-\frac{1}{2})$-th power comes from the Stirling approximation for large k. This roughly explains Theorem 7.1.

The proof is given in Appendix and the result is also verified by simulation in Figure 5.

8 Simulation

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 generated synthetic MAG graphs with varying parameter values to explore how the network properties change as a function of those parameters. We focus on the change of scalar network properties, like diameter and the fraction of nodes in the largest connected component of the graph, as a function of the model parameter values. Second, we also ran simulations with fixed parameter configurations to check other properties of MAG model that we did not theoretically analyze. In this way, we are able to qualitatively compare our model to a 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 parameter. We vary μ, α, f, and n in $M(n, l, \mu, \Theta)$, where α is the first entry of the affinity matrix $\Theta = [\alpha\ \beta; \beta\ \gamma]$ and f indicates a scalar factor of Θ, *i.e.*, $\Theta = f \cdot \Theta_0$ for a constant $\Theta_0 = [\alpha_0\ \beta_0; \beta_0\ \gamma_0]$.

Figure 3 depicts the number of edges, the fraction of nodes in the largest connected component, and the effective diameter (the 90th-percentile distance [28]) of the network as a function μ, α, f, and n for fixed $l = 8$. First, we notice that the growth of network in the number of edges is slower than exponential since the curves on the plot grow sub-linearly in Figure 3(a) with log scaled y-axis. Note that the network size is roughly proportional to $n^2 \left(\mu^2\alpha + 2\mu(1-\mu)\beta + (1-\mu)^2\gamma\right)^l$ from Theorem 3.1. For example, by this formula, the network size is proportional to the l-th power of f, *i.e.*, the eighth power of f in this case. As 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 is, the closer to the straight line the network size curve becomes. For instance, the network size grows by the polynomial function of degree 16 over μ, whereas it grows by degree 2 over n. In 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 Erdös-Rényi random graphs model [14]. The vertical line in the middle of each figure represents the theoretical theshold for the unique giant connected component. As we analyzed, each network contains at least half size of giant connected component at its threshold.

Last, 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 [28, 33].

Furthermore, we also performed simulations where we fix Θ and μ but simultaneously increase both n and l by maintaining 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 MAG model follows densification power law (DPL) and the shrinking diameter properties of real-world networks [28]. Depending on the choice of μ, one can control the rate of densification and the diameter.

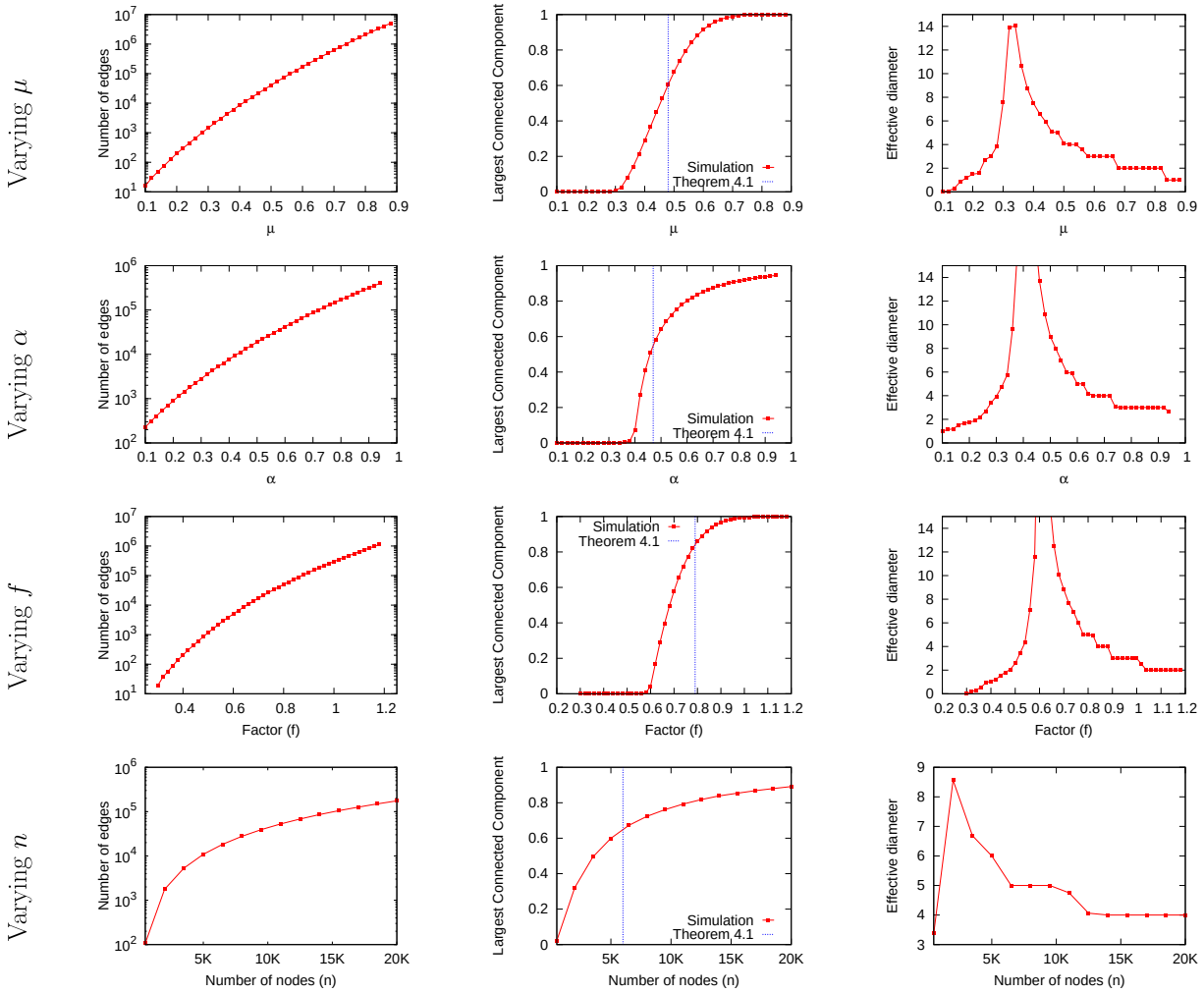
8.2 Degree Distributions

In addition to the network size, connectivity, and diameter, we also examined the degree distributions of MAG graphs empricially. 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 versions of networks and compare their degree distributions.

Figure 5 exhibits the degree distributions of the two types of MAG model. While Figure 5(a) plots the degree distributions of the simplified MAG model $M(n, l, \mu, \Theta)$, Figure 5(b) shows those of the power-law MAG model $M(n, l, \vec{\mu}, \vec{\Theta})$. For each case, the left plot represents the raw form of degree histogram, whereas the right curve plots the *complementary cumulative distribution* (CCDF), which nicely removes the noisy factor. In Figure 5(a), both raw and CCDF versions of distribution look parabolic on the log-log scale, which verifies that $M(n, l, \mu, \Theta)$ has a log-normal degree distribution. On the other hand, in Figure 5(b), both plots exhibit the straight line on the same scale, which indicates that the degree distribution of $M(n, l, \vec{\mu}, \vec{\Theta})$ follows a power-law. All these experimental results agree with our analyses in Section 6 and Section 7.

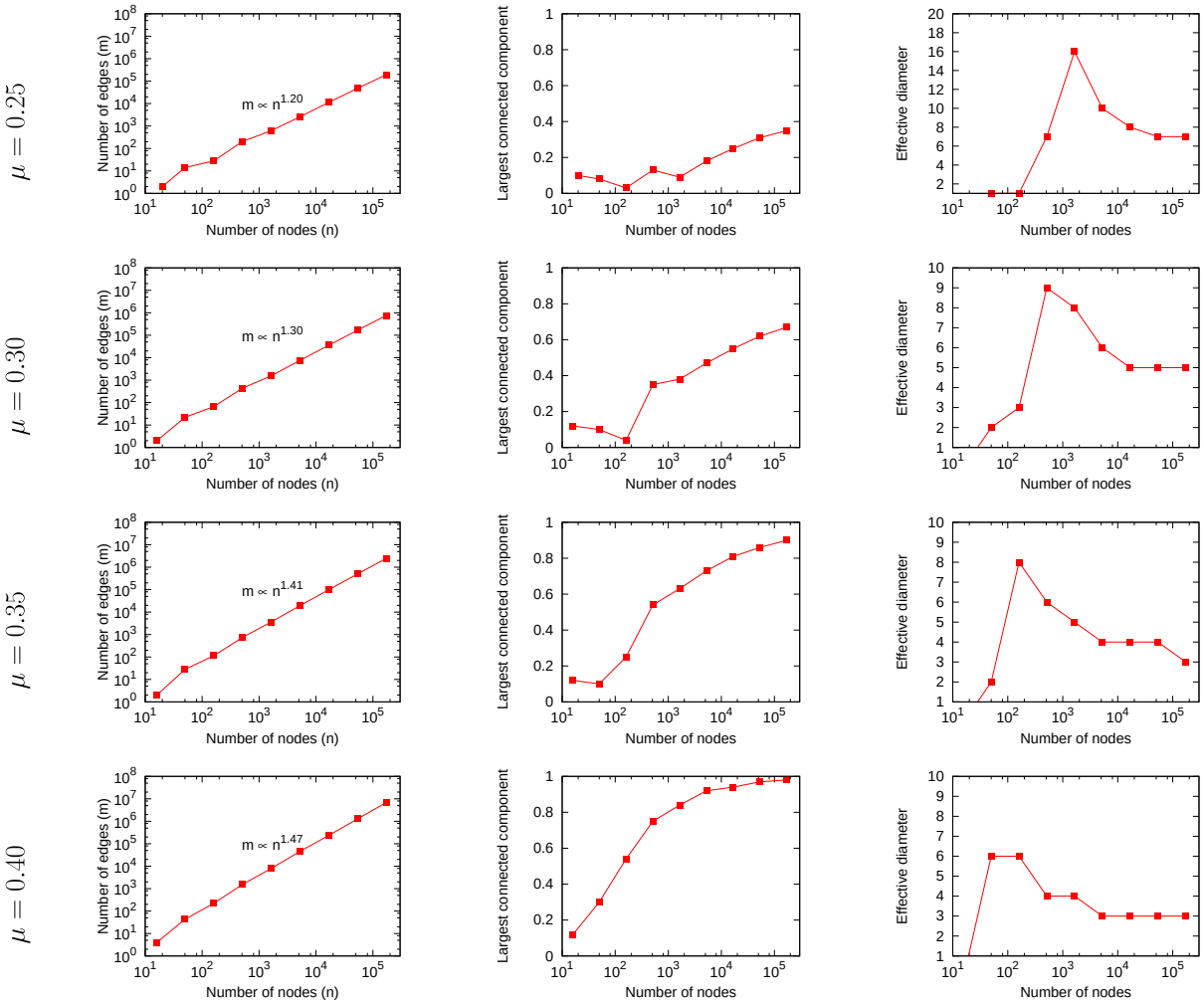
8.3 Comparison to Real-world Networks

Also, we qualitatively compare the structural properties of a specific real-world network and the corresponding MAG model. This leads to interesting questions of how to find optimal MAG model parameters so that synthetic network resembles the given real-world network. The full resolution of these questions lies beyond the scope of the present paper; currently, we searched by brute force over (the relatively small number of) possible MAG parameter settings. We manually selected some parameter settings (for n, l, μ, Θ) to synthesize the simplified MAG model and obtained the properties of $M(n, l, \mu, \Theta)$ to compare the MAG model with a 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 show that many properties of the MAG model exhibit qualitatively similar behavior as real-world networks.



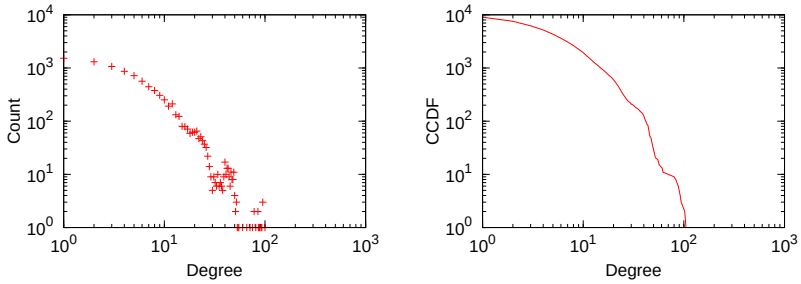
(a) Network size (b) Largest connected component (c) Effective diameter

Figure 3: Structural properties of a simplified MAG model $M(n, l, \mu, \Theta)$ when we fix l and vary a single parameter one by one: μ, α, f, or n. As each parameter increases, in general, the synthetic network becomes denser so that a giant connected component emerges and the diameter decreases to approach a constant.

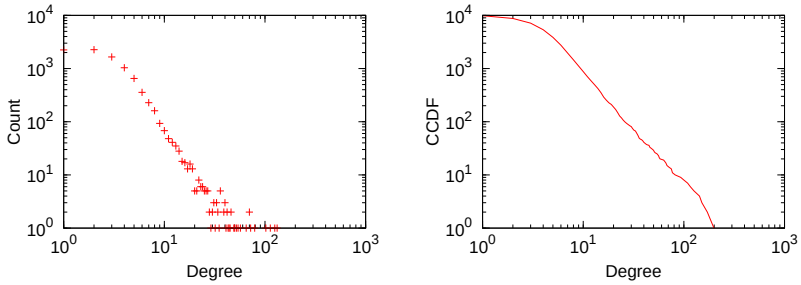


(a) Network size (b) Largest connected component (c) Effective diameter

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 affinity matrix $\Theta = [0.85\ 0.7; 0.7\ 0.15]$ and the ratio $\rho = l/\log n = 0.596$). Observe not only that the relationship between the number of edges and nodes obeys Densification Power Law but also that the diameter begins shrinking after the giant component is formed [33].



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



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

Figure 5: Degree distributions of simplified and power-law version of MAG graphs (see Section 7). We plot both PDF and CCDF of the degree distribution. The simplified version in Figure (a) has parabolic shape on 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.

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 $M(n, l, \mu, \Theta)$, we used $l = 8, \mu = 0.45, \Theta = [0.85\ 0.30; 0.30\ 0.25]$ with the same number of nodes $n = 10,240$. Figure 6(a) and (b) illustrate the following properties of the real-world and the corresponding synthetic network of the simplified MAG model (in the same order of figures).

- *Degree distribution* is a histogram of the number of edges of a node [15].
- *Singular values* indicate the singular values of the adjacency matrix versus their rank [16].
- *Singular vector* represents the distribution of components in the left singular vector associated with the largest singular value [12].
- *Clustering coefficient* represents the degree versus the average (local) clustering coefficient of nodes of a given degree [43].
- *Triad participation* indicates the number of triangles that a node is adjacent to. It measures the transitivity in networks [41].
- *Hop plot* shows the number of reachable pairs of nodes as the number of hops. It sketches how quickly the network expands [40, 27].

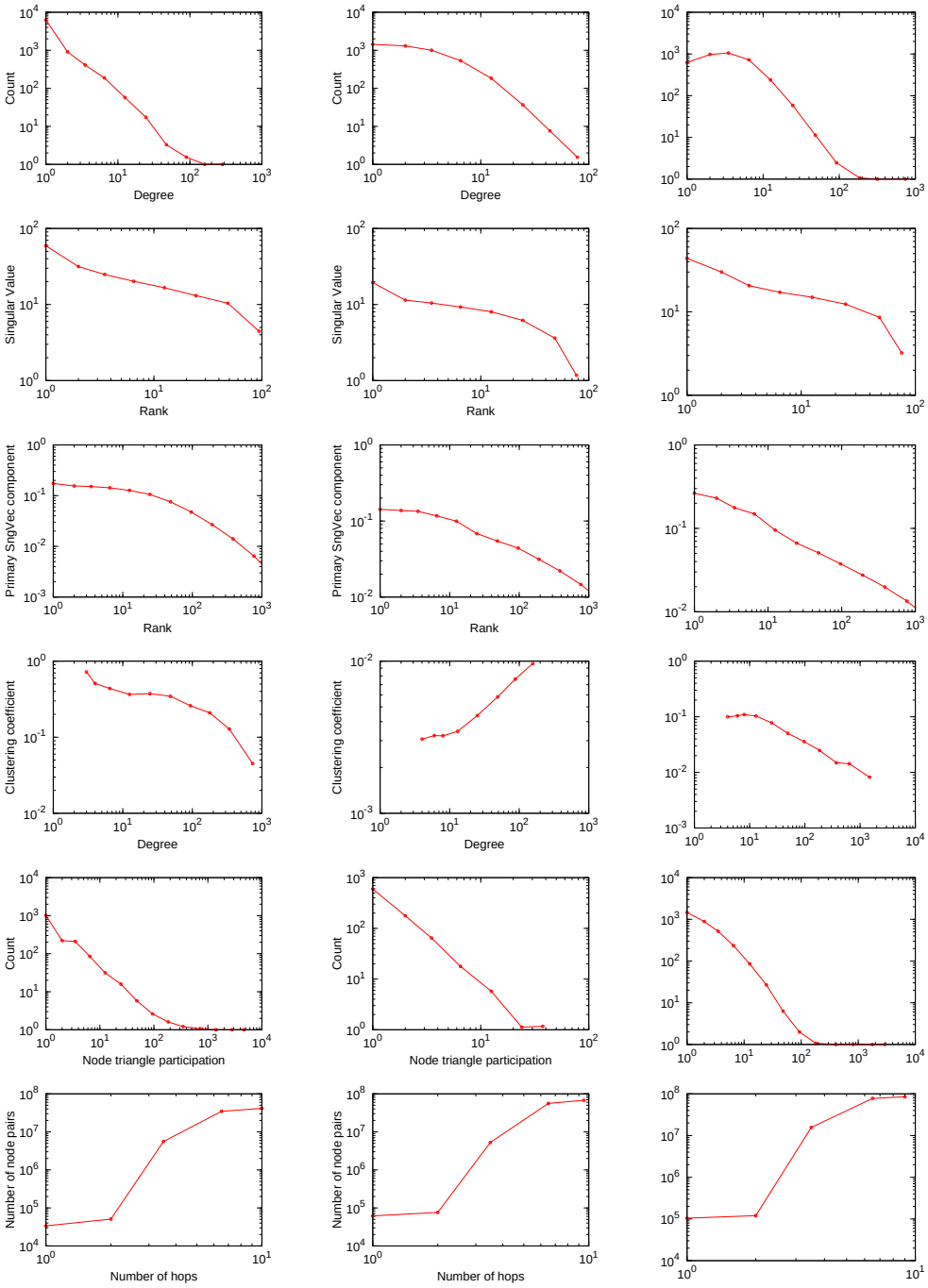
Figure 6 reveals that the plots of properities of MAG model resemble those of Yahoo!-Flickr network. Notice qualitatively similar behavior of nearly all properties between Figure 6(a) and (b). The only property where the simplified MAG model does not match the Yahoo!-Flickr network seems to be the clustering coefficient. As in real-world networks high degree nodes tend to have lower clustering, in the simplified MAG model the situation is reverse – higher degree nodes also tend to have higher clustering. This is due to the fact that for all attributes we use the same affinity matrix Θ which represents only the core-perphery structure ($\alpha > \beta > \gamma$). Thus, the simpified MAG model can only resemble the overall core-periphery shape of real-world networks. However, in the Yahoo!-Flickr network, we can also discover the local clustering effect of homophily and network community formation, which views the network in the opposite way compared to the global core-periphery structure.

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

By this experiment, we can find that MAG model is capable of representing real-world networks. Furthermore, we verify the flexibility of 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 presented the Multiplicative Attribute Graph model for real-world networks which considers the categorical node attributes as well as the affinity of link formation depending on the values of node attributes.



(a) Yahoo!-Flickr network (b) Simplified MAG model (c) General MAG model

Figure 6: The comparison of network properties between real-world Yahoo!-Flickr online social network, a simplified MAG model network, and a general version of MAG model. Except for clustering coefficient, the properties of 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 similar shape to real-world networks in Figure (c).

We introduced the attribute-attribute affinity matrix to represent the affinity of link formation and provide the flexibility in the network structure.

On the other hand, the MAG model is both analytically tractable and statistically interesting. In this paper, we analytically showed 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 unique giant connected component and a small diameter in the MAG model. Furthermore, we mathematically analyzed that the MAG model give rise to networks with either a log-normal or a power-law degree distribution. Finally, we emprically verified our analytical results.

The MAG model is statistically interesting in a sense that it can represent various types of network structure as well as lead a problem that aims to find such structures of the given real-world networks in terms of the MAG model parameters. However, we leave the parameter fitting problem as a venue of the 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.

Acknowledgments

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

References

[1] W. Aiello, F. Chung, and L. Lu. A random graph model for massive graphs. In *STOC '00*, pages 171–180, 2000.

[2] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing. Mixed membership stochastic blockmodels. *JMLR*, 9:1981–2014, 2007.

[3] R. Albert, H. Jeong, and A.-L. Barabási. Diameter of the world-wide web. *Nature*, 401:130–131, 1999.

[4] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[5] A. Blum, H. Chan, and M. Rwebangira. A random-surfer web-graph model. In *ANALCO '06*, 2006.

[6] E. Bodine-Baron, B. Hassibi, and A. Wierman. Distance-dependent kronecker graphs for modeling social networks. *IEEE THEMES*, 2010.

[7] B. Bollobás. The diameter of random graphs. *IEEE Trans. Inform. Theory*, 36(2):285–288, 1990.

[8] B. Bollobás and O. Riordan. Mathematical results on scale-free random graphs. In S. Bornholdt and H. Schuster, editors, *Handbook of Graphs and Networks*, pages 1–37. Wiley, 2003.

[9] A. Bonato, J. Janssen, and P. Pralat. The geometric protean model for on-line social networks. In *WAW '10*, 2010.

[10] 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*, pages 135–144, 2007.

[11] 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.

[12] D. Chakrabarti, Y. Zhan, and C. Faloutsos. R-mat: A recursive model for graph mining. In *SDM '04*, 2004.

[13] C. Cooper and A. Frieze. A general model of web graphs. *RSA*, 22(3):311–335, 2003.

[14] P. Erdős and A. Rényi. On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Acadamy of Science*, 5:17–67, 1960.

[15] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *SIGCOMM '99*, pages 251–262, 1999.

[16] I. J. Farkas, I. Derényi, A.-L. Barabási, and T. Vicsek. Spectra of ıreal-world graphs: Beyond the semicircle law. *Phys. Rev. E*, 64(2):026704, Jul 2001.

[17] A. D. Flaxman, A. M. Frieze, and J. Vera. A geometric preferential attachment model of networks. In *WAW '04*, pages 44–55, 2004.

[18] P. Hoff and A. Raftery. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 2002.

[19] M. Kim and J. Leskovec. Multiplicative attribute graph model of real-world networks. In *WAW '10*, 2010.

[20] M. Kim and J. Leskovec. Network completion problem: Inferring missing nodes and edges in networks. In *SDM '11*, 2011.

[21] V. Klee and D. Larmann. Diameters of random graphs. *Canad. J. Math.*, 33:618–640, 1981.

[22] J. M. Kleinberg. Navigation in a small world. *Nature*, 406(6798), August 2000.

[23] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. In *FOCS '00*, page 57, 2000.

[24] S. Lattanzi and D. Sivakumar. Affiliation networks. In *STOC '09*, pages 427–434, 2009.

[25] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani. Kronecker Graphs: An Approach to Modeling Networks. *JMRL*, 11:985–1042, 2010.

[26] J. Leskovec, D. Chakrabarti, J. M. Kleinberg, and C. Faloutsos. Realistic, mathematically tractable graph generation and evolution, using kronecker multiplication. In *PKDD '05*, pages 133–145, 2005.

[27] J. Leskovec and C. Faloutsos. Scalable modeling of real graphs using kronecker multiplication. In *ICML '07*, 2007.

[28] J. Leskovec, J. M. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *KDD '05*, pages 177–187, 2005.

[29] 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.

[30] 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):29–123, 2009.

[31] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. Geographic routing in social networks. *PNAS*, 102(33):11623–11628, Aug 2005.

[32] M. Mahdian and Y. Xu. Stochastic kronecker graphs. In *WAW '07*, 2007.

[33] M. McGlohon, L. Akoglu, and C. Faloutsos. Weighted graphs and disconnected components: patterns and a generator. In *KDD*, pages 524–532, 2008.

[34] M. McPherson. An ecology of affiliation. *American Sociological Review*, 48(4):519–532, 1983.

[35] M. J. McPherson and J. R. Ranger-Moore. Evolution on a dancing landscape: Organizations and networks in dynamic blau space. *Social Forces*, 70(1):19–42, 1991.

[36] S. Milgram. The small-world problem. *Psychology Today*, 2:60–67, 1967.

[37] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1:226–251, 2004.

[38] M. Mitzenmacher. Editorial: The future of power law research. *Internet Mathematics*, 2:525–534, 2006.

[39] G. Palla, L. Lovász, and T. Vicsek. Multifractal network generator. *PNAS*, 107(17):7640–7645, 2010.

[40] 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.

[41] C. E. Tsourakakis. Fast counting of triangles in large real networks without counting: Algorithms and laws. *ICDM '08*, 2008.

[42] S. Wasserman and P. Pattison. Logit models and logistic regressions for social networks. *Psychometrika*, 60:401–425, 1996.

[43] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[44] P. M. Weichsel. The kronecker product of graphs. *American Mathematical Society*, 13(1):37–52, 1962.

[45] S. J. Young and E. R. Scheinerman. *Random Dot Product Graph Models for Social Networks*, volume 4863 of *Lecture Notes in Computer Science*. Springer Berlin, 2007.

A Appendix: The Number of Edges

Proof of Lemma 3.2: Let N_{uv}^0 be the number of attributes that take 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 value 1 in both u and v. Then, $N_{uv}^0, N_{uv}^1 \geq 0$ and $N_{uv}^0 + N_{uv}^1 \leq l$ as l indicates the number of attributes in each node.

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

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

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, *i.e.*, $u \in W_i$. Since u and v can share value 0 only for the attributes where u already takes value 0, N_{uv}^0 equivalently represents the number of heads in i coin flips with probabiltiy μ. Therefore, N_{uv}^0 follows $Bin(i,\mu)$. Similarly, N_{uv}^1 follows $Bin(l-i, 1-\mu)$. Hence, their conditional joint probability is

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

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

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

■

Proof of Lemma 3.3: By Lemma 3.2 and the linearity of expectation, we sum this conditional probability over all nodes and result in the expectation of the degree given the weight of node u. ■

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

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

■

Proof of Corollary 3.3.1: Suppose that $l=\left(\epsilon-\frac{1}{\log\zeta}\right)\log n$ 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 $\Theta\left(n^2\zeta^l\right)$. Note that $\log\zeta<0$ since $\zeta<1$. Therefore, the expected number of edges is

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

■

Proof of Corollary 3.3.2: Under the situation that $l\in o(\log n)$, the expected number of edges is

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

■

B Appendix: Connectivity

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 $\mathbb{E}\left[P\left[u, v^{(j)}\right] | v \in W_i\right] = \mathbb{E}\left[P\left[u, v^{(j)}\right] | v^{(j)} \in W_j\right]$, because each $v^{(j)}$ can be generated by $\binom{j}{i}$ different $a(v)$ sets with the same probability. Therefore,

$$\mathbb{E}\left[P\left[u, v\right] | v \in W_j\right] = \mathbb{E}\left[\mathbb{E}\left[P\left[u, v^{(j)}\right] | v \in W_i\right]\right] \geq \mathbb{E}\left[\mathbb{E}\left[P\left[u, v\right] | v \in W_i\right]\right] = \mathbb{E}\left[P\left[u, v\right] | v \in W_i\right] .$$

■

Next theorem plays a key role in proving Theorem 4.1 as well as Theorem 4.2.

Theorem B.1 *Let $|S_j| \in \Theta(n)$ and $\mathbb{E}\left[P\left[u, V\backslash u\right] | u \in W_j\right] \geq c \log n$ as $n \to \infty$ for some j and sufficiently large c. Then, S_j is connected with high probability 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

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

for distinct u and v. By Theorem 4.3,

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

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

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

as $n \to \infty$. Therefore, S_j is connected with high probability. ■

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

Lemma B.2 *$|S_{\mu l}| \geq \frac{n}{2} - o(n)$ with high probability as $n \to \infty$.*

Proof: By Central Limit Theorem, $\frac{|u|-\mu l}{\sqrt{l\mu(1-\mu)}} \sim N(0,1)$ as $n \to \infty$, *i.e.*, $l \to \infty$. Therefore, $P(|u| \geq \mu l)$ is at least $\frac{1}{2} - o(1)$ so $|S_{\mu l}| \geq \frac{n}{2} - o(n)$ with high probability as $n \to \infty$. ■

Lemma B.3 *$|S_{\mu l+l^{1/6}}| \in \Theta(n)$ with high probability as $n \to \infty$.*

Proof: By Central Limit Theorem mentioned in Lemma B.2,

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

as $l \to \infty$ where $\Phi(z)$ represents the cdf of the standard normal distribution.
Since $P(|u| \geq \mu l + l^{1/6})$ is still at least $\frac{1}{2} - o(1)$, the size of $S_{\mu l+l^{1/6}}$ is $\Theta(n)$ with high probability as $l \to \infty$, *i.e.*, $n \to \infty$. ■

Lemma B.4 *$|S_{\mu l+l^{2/3}}| \in o(n)$ with high probability as $n \to \infty$.*

Proof: By Chernoff bound, $P(|u| \geq \mu l + l^{2/3})$ is $o(1)$ as $l \to \infty$, thus $|S_{\mu l+l^{2/3}}|$ is $o(n)$ with high probability as $n \to \infty$. ■

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

Proof of Theorem 4.1:

(Existence). First, if $\left[(\mu\alpha + (1-\mu)\beta)^{\mu} (\mu\beta + (1-\mu)\gamma)^{1-\mu}\right]^{\rho} > \frac{1}{2}$, then by Lemma 3.3,

$$\mathbb{E}\left[P\left[u, V \backslash u\right] | u \in W_{\mu l}\right] \approx \left[2\left[(\mu\alpha + (1-\mu)\beta)^{\mu} (\mu\beta + (1-\mu)\gamma)^{1-\mu}\right]^{\rho}\right]^{\log n} = (1+\epsilon)^{\log n} > c \log n$$

for some constant $\epsilon > 0$ and $c > 0$. Since $|S_{\mu l}| \in \Theta(n)$ by Lemma B.2, $S_{\mu l}$ is connected with high probability as $n \to \infty$ by Theorem B.1. In other words, we are able to extract out a connected component of size at least $\frac{n}{2} - o(n)$.

Second, when $\left[(\mu\alpha + (1-\mu)\beta)^{\mu} (\mu\beta + (1-\mu)\gamma)^{1-\mu}\right]^{\rho} = \frac{1}{2}$, we can apply the same argument for

$S_{\mu l+l^{1/6}}$. Because $|S_{\mu l+l^{1/6}}| \in \Theta(n)$ by Lemma B.3,

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

which is also greater than $c\log n$ as $n\to\infty$ for some constant $\epsilon'>0$. Thus, $S_{\mu l+l^{1/6}}$ is connected with high probability by Theorem B.1.

Last, on the contrary, when $\left[\left(\mu\alpha+(1-\mu)\beta\right)^{\mu}\left(\mu\beta+(1-\mu)\gamma\right)^{1-\mu}\right]^{\rho}<\frac{1}{2}$, for $u\in W_{\mu l+l^{2/3}}$,

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

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 $\mu l+l^{2/3}$ is $o(1)$. However, since $S_{\mu l+l^{2/3}}$ is $o(n)$ by Lemma B.4, $n-o(n)$ nodes have less than $\mu l+l^{2/3}$ weights. Hence, most of $n-o(n)$ nodes are isolated so that the size of the largest component cannot be $\Theta(n)$.

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

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

Since $V\backslash H\subset V\backslash S_{\mu l}$,

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

holds for every $u\in V$.

Suppose that another connected component H' also contains $\Theta(n)$ nodes. We will show the condtradiction if H and H' are not connected with high probability as $n \to \infty$. To see $\mathbb{E}\left[P\left[H, H'\right]\right]$,

$$\begin{aligned}\mathbb{E}\left[P\left[H, H'\right]\right] &= |H'| \cdot \mathbb{E}\left[P\left[u, H\right] | u \in H'\right] \\ &\geq \frac{|H'| \cdot |S_{\mu l}|}{n - |S_{\mu l}|} \mathbb{E}\left[P\left[u, V \backslash S_{\mu l}\right] | u \in H'\right] \\ &\geq \frac{|H'| \cdot |S_{\mu l}|}{n - |S_{\mu l}|} \mathbb{E}\left[P\left[u, H'\right] | u \in H'\right] \qquad (\because H' \subset V \backslash H \subset V \backslash S_{\mu l}) .\end{aligned}$$

However, $\mathbb{E}\left[P\left[u, H'\right] | u \in H'\right] \in \Omega(1)$. Otherwise, since the probability that $u \in H'$ is connected to H' is not greater than $\mathbb{E}\left[P\left[u, H'\right] | u \in H'\right]$ by Markov Inequality, u is disconnected from H' with high probability as $n \to \infty$. H' thus includes at least one isolated node with high probability as $n \to \infty$. This is contradiction to the connectedness of H'.

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

Next, the proofs for the connectedness follow. Before the main proof, we present some necessary lemmas and prove them.

Lemma B.5 *$\left(\frac{\mu}{x}\right)^x \left(\frac{1-\mu}{1-x}\right)^{1-x}$ is a monotonically increasing function of x over $(0, \mu)$.*

Proof: Let $f(x)$ be the log-value of the given function, ***i.e.,***

$$f(x) = x\left(\log \mu - \log x\right) + (1 - x)\left(\log\left(1 - \mu\right) - \log\left(1 - x\right)\right) .$$

To take the derivative of $f(x)$,

$$f'(x) = \left(\log \mu - \log x\right) + \left(\log\left(1 - x\right) - \log\left(1 - \mu\right)\right) .$$

Since $x < \mu$ and $1 - \mu < 1 - x$, $f'(x) > 0$ where $0 < x < \mu$. This implies that $f(x)$ is strictly increasing, so the given function is also strictly increasing over $(0, \mu)$. ■

Lemma B.6 *If $(1-\mu)^\rho \geq \frac{1}{2}$, then $\frac{V_{\min}}{l} \to 0$ with high probability as $n \to \infty$. Otherwise, if $(1-\mu)^\rho < \frac{1}{2}$, $\frac{V_{\min}}{l} \to \nu$ with high probability 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 $n(1-\mu)^\rho \geq 1$ by defition. Then, the probability that $|W_i| = 0$ is at most $\exp(-\frac{1}{2}\mathbb{E}\left[|W_i|\right])$ by Chernoff bound. However, for fixed μ,

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

Therefore, by Chernoff bound, $P(|W_1| = 0) \to 0$ as $l \to \infty$. This implies that $V_{\min}$ is $o(l)$ with high probability as $n \to \infty$.

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

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

Since $\left(\frac{\mu}{x}\right)^x\left(\frac{1-\mu}{1-x}\right)^{1-x}$ is a increasing function of x over $(0,\mu)$ by Lemma B.5,

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

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

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

exponentially increases as l increases. By Chernoff bound, $|W_{(\nu+\epsilon)l}|$ is not zero with high probability as $l \to \infty$, *i.e.*, $n \to \infty$.

In a similar way, $\mathbb{E}\left[|W_{(\nu-\epsilon)l}|\right] = \frac{(1-\epsilon')^l}{\sqrt{2\pi l(\nu-\epsilon)(1-(\nu-\epsilon))}}$ exponentially decreases as l increases. Since $\mathbb{E}\left[|W_i|\right] \geq \mathbb{E}\left[|W_j|\right]$ if $\mu l \geq i \geq j$, the expected number of nodes with at most weight $(\nu-\epsilon)l$ is less than $(\nu-\epsilon)l\,\mathbb{E}\left[|W_{(\nu-\epsilon)l}|\right]$ and its value goes to zero as $l \to \infty$. Hence, by Chernoff bound, there exists no node of the weight less than $(\nu-\epsilon)l$ with high probability as $n \to \infty$.

To sum up, $\frac{V_{\min}}{l}$ goes to ν with high probability as $l \to \infty$, *i.e.*, $n \to \infty$. ■

Using the above lemmas, we show the condition that the network is connected.

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

If $\left[(\mu\alpha+(1-\mu)\beta)^t\left(\mu\beta+(1-\mu)\gamma\right)^{1-t}\right]^\rho > \frac{1}{2}$, by Lemma 3.3,

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

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 $\Theta(n)$, $S_{V_{\min}}$ is connected with high probability as $n \to \infty$ by Theorem B.1. Equivalently, the entire network is also connected with high probability $n \to \infty$.

On the other hand, when $(\mu\alpha+(1-\mu)\beta)^{\frac{V_{\min}}{\log n}}(\mu\beta+(1-\mu)\gamma)^{\frac{l-V_{\min}}{\log n}} < \frac{1}{2}$, the expected degree of a node with $|V_{\min}|$ weight is $o(1)$ because from the above relationship $\mathbb{E}\left[P\left[u,V\backslash u\right]|u\in W_{V_{\min}}\right] \approx (1-\epsilon')^{\log n}$ for some $\epsilon'>0$. Thus, in this case, some node in $W_{V_{\min}}$ is isolated with high probability so the network is disconnected. ■

C Appendix: Diameter

Theorem C.1 *[7, 21] For an Erdös-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 5.3: Let A^G and A^H be the probabilistic adjacency matrix 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 with high probability, then so does G. It 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, as the edges created in the first step already result in the constant diameter with high probability, G has a constant diameter.

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

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

for some $\epsilon>0$.

Since this value goes to infinity as $n\to\infty$, so does $\mathbb{E}\left[W_{\lambda l}\right]$. Therefore, by Chernoff bound, $|W_{\lambda l}| \geq c\mathbb{E}\left[W_{\lambda l}\right]$ with high probability as $n\to\infty$ for some constant c. Then,

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

By Theorem C.1, an Erdös-Rényi random graph $G(|S_{\lambda l}|, \frac{c(1+\epsilon)^{\log n}}{|S_{\lambda l}|\sqrt{2\pi l\lambda(1-\lambda)}})$ has a diameter of at most $\left(1+\frac{\ln 2}{\epsilon}\right)$ with high probability as $n\to\infty$. Thus, the diamters of $G(|S_{\lambda l}|, \beta^{\lambda l}\gamma^{(1-\lambda)l})$ as well as $S_{\lambda l}$ are also bounded by a constant with high probability as $n\to\infty$. ■

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

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

By Centeral Limit Theorem, $\sum_{j=\lambda l}^{l}\binom{l}{j}\lambda^j(1-\lambda)^{l-j}$ converges to $\frac{1}{2}$ as $l \to \infty$. Therefore, $P[u, S_{\lambda l}]$ is greater than $c\log n$ for a constant c, and then, by Chernoff bound, u is directly connected to $S_{\lambda l}$ with high probability as $n \to \infty$. ■

D Appendix: Degree Distribution

Theorem D.1 *[45]* $P\left(deg(u)=k\right) = \int_{u\in V}\binom{n-1}{k}\left(\mathbb{E}\left[P\left[u,v\right]\right]\right)^k\left(1-\mathbb{E}\left[P\left[u,v\right]\right]\right)^{n-1-k}du$.

Corollary D.1.1 *For* $E_j = (\mu\alpha+(1-\mu)\beta)^j\left(\mu\beta+(1-\mu)\gamma\right)^{l-j}$,
the probability of degree k *in* $M(n,l,\mu,\Theta)$ *is* $p_k = \sum_{j=0}^{l}\binom{l}{j}\mu^j(1-\mu)^{l-j}\binom{n-1}{k}E_j^k\left(1-E_j\right)^{n-1-k}$.

Proof: To reformulate Theorem D.1,

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

Therefore, by applying Lemma 3.2, we obtain the desired formula. ■

Proof of Theorem 6.1: To reduce the space, we begin by defining some notations as follow:

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

By Corollary D.1.1, we can restate p_k as $\sum_{j=0}^{l}g_j(k)$.

If most of those terms turn out to be insignificant under our assumptions, the probability p_k can be approximately proportional to one or few dominant terms. In this case, what we need to do is thus to seek for j that maximizes $g_j(k) = \binom{l}{j}\mu^j(1-\mu)^{l-j}f_j(k)$ and find its approximate formula.

We start with the approximation of $f_j(k)$. For large n and k, by Stirling approximation,

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

However, the expected degree of maximum weight node is $O(n\left(\mu\alpha+(1-\mu)\beta\right)^l)$, so is the expected maximum degree. k is thus $o(n)$ with high probability as $n\to\infty$, *i.e.*, $l\to\infty$.

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

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

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

for some constant C. When $k = nx^\tau y^{l-\tau}$ for $\tau \geq \mu l$ and $R = \frac{x}{y}$,

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

Using $(j-\mu l)^2 = (j-\tau)^2 + (\tau-\mu l)^2 + 2(j-\tau)(\tau-\mu l)$,

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

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

Considering $g_j(k)$ as a function of j, not k, now we find j that maximizes $g_j(k)$ for $k = nx^\tau y^{l-\tau}$. However, the median weight is approximately equal to μl by Central Limit Theorem. If we focus on the higher half degrees. we can thus let $\tau \geq \mu l$. In this case, since $\left[(\mu\alpha+(1-\mu)\beta)^\mu \left(\mu\beta+(1-\mu)\gamma\right)^{1-\mu}\right]^\rho > \frac{1}{2}$,

$$\therefore \quad k \geq \left[(\mu\alpha+(1-\mu)\beta)^\mu \left(\mu\beta+(1-\mu)\gamma\right)^{1-\mu}\right]^\rho \in \Omega(l) .$$

If we differentiate $\ln g_j(k)$ over j,

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

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

Furthermore, since $|\frac{j-\tau}{2l\mu(1-\mu)}| \ll k \ln R$ as $n \to \infty$, the first quadratic term $\frac{(j-\tau)^2}{2l\mu(1-\mu)}$ in $\ln g_j(k)$ is negligible. As a result, when R is practical (close to $1.6 \sim 3$), $\ln g_{\tau+\Delta}$ would be at most $(\Theta(-k|\Delta|) - \ln g_\tau)$ for $\Delta \geq 1$. After all, g_τ effectively dominates the probability p_k, *i.e.*, $\ln p_k$ is roughly proportional to $\ln g_\tau$. By assigning $\tau = \frac{\ln k - \ln ny^l}{\ln R}$, we obtain

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

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

E Appendix: Power-law Distribution

Proof of Lemma 7.2: Since a_i's are independently distributed Bernoulli random variables, Lemma 7.2 holds. ■

Proof of Lemma 7.3: Let's define $P_j(u,v)$ as the edge probability between u and v when considering only up to the j-th attribute, *i.e.*,

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

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

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

When $l = 1$, it is trivially true by Lemma 3.2. When $l > 1$, suppose that the above formula holds for $l = 1, 2, \cdots, k$. Since $P_{k+1}(u,v) = P_k(u,v)\Theta_{k+1}\left[a_{k+1}(u), a_{k+1}(v)\right]$,

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

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

Proof of Theorem 7.1: Before the main argument, we need to define the ordered probability mass of attribute vectors as $p_{(j)}$ for $j = 1, 2, \cdots, 2^l$. 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 $l = 2$, the ordered probability mass is $p_{(1)} = 0.1$, $p_{(2)} = 0.2$, $p_{(3)} = 0.3$, and $p_{(4)} = 0.4$.

Then, by Theorem D.1, we can express the probability of degree k, p_k, as follows:

$$p_k = \binom{n-1}{k} \sum_{j=1}^{2^l} p_{(j)} (E_j)^k (1-E_j)^{n-1-k} \tag{2}$$

where E_j denotes the average edge probability of the node which has the attribute vector corresponding to $p_{(j)}$. If $p_{(j)}$'s and E_j's are configured so that few terms dominate the probability, we may approximate p_k as $\binom{n-1}{k} p_{(\tau)} (E_\tau)^k (1-E_\tau)^{n-1-k}$ for $\tau = \arg\max_j p_{(j)} (E_j)^k (1-E_j)^{n-1-k}$. 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 condtion that $p_{(j)} \propto E_j^{-\delta}$ for a constant δ. Then, the j-th term is

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

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

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

is $o(1)$ for $\Delta \geq 1$ as $n \to \infty$. Therefore, the τ-th term dominates the Equation (2).

Next, by the Stirling approximation with above conditions,

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

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

Last, we prove that the two conditions for the power-law degree distribution are simultaneously feasible by providing an example configuration.

If every $p_{(j)}$ is 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 that $p_{(j)} \propto (E_j)^{-\delta}$ by Lemma 7.2 and Lemma 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 μ_i's and Θ_i's independently, the sufficient condition for the power law degree distribution is feasible. ■