Multiplicative Attribute Graph Model of Real-World Networks

Myunghwan Kim  Jure Leskovec
Stanford University

Abstract

Large scale real-world network data, such as social networks, Internet and Web graphs, are ubiquitous. The study of such social and information networks seeks to find patterns and explain their emergence through tractable models. In most networks, especially in social networks, nodes have a rich set of attributes (e.g., age, gender) associated with them. However, many existing network models focus on modeling the network structure while ignoring the features of the nodes.

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 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 similarities. This 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 graphs with a constant diameter. We analyze the degree distribution to show that the model can produce networks with either log-normal or power-law degree distribution depending on certain conditions.

1 Introduction

With the emergence of the Web, large online social computing applications have become ubiquitous, which then 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 is then to study real-world networks, with an emphasis on finding patterns of connectivity and explaining 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? 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 [13, 10], local clustering of edges [35], small diameter [3, 23], navigability [30, 17], emergence of community structure [24], and so on.

In parallel, there have been efforts to find the network formation mechanisms that naturally generate networks with such 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 network models where connectivity patterns observed in the real-world naturally emerge from the model. The prime example of this line of work is the Preferential Attachment model with its many variants [4, 11, 8, 9], which specifies a simple but very natural edge creation mechanism that in the limit leads to networks with power-law degree distributions. Other examples of models of similar flavor include the Copying Model [18],
the Small-world model [35], Geometric Random Graphs [14], the Forest Fire model [23], the Random surfer model [5], and models of bipartite affiliation networks [19]. On the other hand, in statistics, machine learning and traditional social network analysis, a different approach to modeling network data has emerged. Here 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 [34], the Stochastic Block Model [2] and the Latent Space Model [15].

“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 that are also 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 be able to model heterogeneities and differences between individual nodes.

On the contrary, the “statistical” models are normally mathematically 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 attributes of nodes and edges.

Although models of network structure or formation are seldom both analytically tractable and statistically interesting, an example of a model satisfying both features is the Kronecker graphs model [21], which is based on the recursive tensor product of small graph adjacency matrices. Kronecker graphs are analytically tractable in a sense that one can analyze global structural properties of networks that emerge from the model [26]. In addition, this model is statistically meaningful because there exists an efficient parameter estimation technique based on maximum likelihood [22]. It is empirically shown that with only four parameters Kronecker graphs quite accurately model the global structural properties of real world graphs 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 are 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 mathematically tractable but statistically impractical or statistically very powerful but do not lend themselves to mathematical analysis.

Here 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) and as well as those reflecting 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 the homophily as well as the heterophily that both naturally occur in social networks.

Next we proceed by formulating the model and show that it is both statistically interesting and mathematically tractable. In the sections that follow we demonstrate our mathematical results. First, we examine the number of edges and shows that our model naturally obeys the Densification Power Law [23]. Second, we examine the connectivity of MAG model, which includes the conditions not only when the model contains a giant connected component but also when it becomes connected. Third, we show that the diameter of the MAG model remains small even though the number of nodes is very large. Fourth, we demonstrate that degree distribution of networks that emerge from the MAG model follows a log-normal distributions. Furthermore, we describe a more general version of the model that can capture different degree distributions. We view this as particularly interesting because there has been a long-standing debate about how to distinguish power-law distributions from log-normal distributions in empirical data [31, 32] and what implications this would make for real-world networks. Also, this indicates that our model is flexible in a sense that networks with very different properties emerge depending on the parameter configuration. Last, we perform simulation experiments to further illustrate the properties of the MAG model. 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

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. 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 would like our model to be able to account for both the homophily of certain features as well as the heterophily of the others.
More precisely, we to associate each feature $i$ (i.e., $i$-th question) with an attribute-attribute similarity matrix $\Theta_i$. For the above binary example, each $\Theta_i$ should be $2 \times 2$ matrix. The entries of matrix $\Theta_i$ represent the probability of an edge given the values of the $i$-th attribute of both nodes, i.e., attribute values act as “selectors” of an appropriate cell of $\Theta_i$. Thus, if the attribute reflects homophily, the corresponding matrix $\Theta_i$ would have large values on the diagonal (i.e., the probability of the edge is high when the nodes’ answers match), whereas if the attribute represents heterophily the off-diagonal values of $\Theta_i$ would be high (i.e., the probability of the link is high when nodes gave different answers to the same question). The top of Figure 1 illustrates the concept of node attributes acting as selectors of entries of matrices $\Theta_i$.

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 $i = 1, 2, \ldots, l$. We also have $l$ matrices, each of which is $\Theta_i \in d_i \times d_i$. Each entry of $\Theta_i$ is a probability, i.e., a real value between 0 and 1. Then, the probability of an edge $(u, v)$, $P[u, v]$, is defined as the multiplication of probabilities corresponding to individual attributes, i.e.,

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

(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 $\Theta_i$. 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 of $u$ to these questions. $\Theta_i$ then reflects the marginal edge probability over the $i$-th answers of for a pair of nodes. That is, the answers of nodes $u$ and $v$ on a question $i$ select an entry of matrix $\Theta_i$, i.e., $u$’s answer selects a row and $v$’s answer selects a column. One can thus think of matrices $\Theta_i$’s as the attribute-attribute similarity matrices. Assuming that the questions are appropriately chosen so that answers should be independent each other, the product over the entries of matrices $\Theta_i$ results in the probability of the edge between $u$ and $v$.

The choice of multiplicatively combining entries of $\Theta_i$ is very natural. In particular, the social network literature defines a concept of Blau-spaces [28, 29] where socio-demographic properties 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).

The proposed model is also mathematically 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 to estimate the values of matrices $\Theta_i$? How to infer the attributes of unobserved nodes? Or, given a network, how to estimate both the node attributes and the matrices $\Theta_i$. For example, an expectation maximization (EM) based framework could be used to estimate the node attributes as well as the attribute-attribute similarity matrices. However, we leave the questions of parameter estimation for the future work and rather focus on the analysis of the structural properties of the networks that emerge from the MAG model.

\footnote{Note that there is no condition for $\Theta_i$ to be stochastic, we only require each entry of $\Theta_i$ to be on interval (0, 1).}
Simplified version of the model. Next we delineate a simplified version of the model that we will then mathematically analyze in the further sections of the paper. First, while the general MAG model applies to directed networks, we will consider undirected version of the model by requiring each $\Theta_i$ to be symmetric. Second, we will assume binary node attributes and thus matrices $\Theta_i$ have 2 rows and 2 columns. Third, to further reduce the number of parameters, we will also assume that the similarity matrices for all attributes are the same, i.e., $\Theta_i = \Theta$ for all $i$. This means that $\Theta = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix}$, i.e., $\Theta[1, 1] = \alpha, \Theta[1, 0] = \Theta[0, 1] = \beta$, and $\Theta[0, 0] = \gamma$ for $0 \leq \alpha, \beta, \gamma \leq 1$. Furthermore, all our results will hold for $\alpha > \beta > \gamma$. As we show later, the assumption $\alpha > \beta > \gamma$ is very natural since most large real-world networks have a common structure [20].

Last, we will 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 $\mu$. That is, we use an i.i.d. Bernoulli distribution parameterized by $\mu$ to model attribute vectors where the probability that the $i$-th attribute of a node $u$ takes value 1 is $P(a_i(u) = 1) = \mu$ for $i = 1, \ldots, l$ 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, $\mu$ is the probability that an attribute takes a value of 1, and $\Theta = [\alpha \beta; \beta \gamma]$ specifies the attribute-attribute similarity 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 $\rho$.

Connections to other models. We note that our model belongs to a general class of latent space network models, where it is assumed that 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 [15] 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 [37], the linking probability depends on the inner product between the vectors associated with node positions. Furthermore, recently introduced Multifractal network generator [33] can also be viewed as a special case of MAG model where the node attribute distribution and the similarity matrix are equal for every attribute.

The MAG model generalizes the Kronecker graphs model [20] in a subtle way. The Kronecker graphs model takes a small (usually $2 \times 2$) initiator matrix $K$ and tensor-powers it $l$ times to obtain a matrix $G$ of size $2^l \times 2^l$, simply interpreted as the stochastic graph adjacency matrix. One can think of a Kronecker graph model as a special case of the MAG model.

Proposition 2.1 A Kronecker graph $G$ on $2^l$ nodes with a $2 \times 2$ initiator matrix $K$ is equivalent to the following MAG graph $M$: Let us number the nodes of $M$ as $0, \ldots, 2^l - 1$. Let the binary attribute vector of a node $u$ of $M$ be simply 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 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 simply 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 $0, \ldots, 2^l - 1$).

\[2\text{Throughout the paper, } \log(\cdot) \text{ indicates } \log_2(\cdot) \text{ unless explicitly specified as } \ln(\cdot).\]
Second, building on this correspondence between Kronecker and MAG graphs, we also note that the estimates of the Kronecker parameter matrix $K$ nicely transfer to matrix $\Theta$ of MAG model. For example, Kronecker parameter 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 $K = [\alpha = 0.99, \beta = 0.53, \gamma = 0.13]$ [22]. Thus, in the rest of the paper, we will consider the above values of $K$ as the typical values that the matrix $\Theta$ would normally take. In this respect, the assumption that $\alpha > \beta > \gamma$ appears as very natural.

Furthermore, the fact that most large real-world networks satisfy $\alpha > \beta > \gamma$ tells us that such networks have an onion-like “core-periphery” structure [24][20]. In other words, the network is composed from denser internet connectivity, while the global network structure of the Epinions online social network is captured by $K$. A more edges are likely to appear between nodes which share 1’s on more attributes and these nodes form the core of the network. Since more edges appear between pairs of nodes with attribute combination “1–0” than between those with “0–0”, there are more edges between the core and the periphery nodes (edges “1–0”) than between the nodes of the periphery themselves (edges “0–0”).

In the following sections, we analyze on the properties of the MAG model. We focus mostly on the simplified version. In each section state the main theorem and give 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 is able not only to validate the assumption, $l = \rho \log n$, but also to derive a substantial social network property with regard to graph density.

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

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

This is divided into two different 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 for convenience. 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 1’s in its attribute vectors, and denote it as $|u|$, i.e., $|u| = \sum_{i=1}^{l} 1 \{a_i(u)=1\}$ where $1 \{ \cdot \}$ 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 $j = 0, 1, \ldots, l$. 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, $S_j = \bigcup_{i=j}^{l} W_i$.

To complete the proof of Theorem 3.1, using the definition of the simplified MAG model and some algebra, 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 $u \in V$, $\mathbb{E}[\deg(u)|u \in W_i] = (n-1) (\mu \alpha + (1-\mu)\beta)^i (\mu \beta + (1-\mu)\gamma)^{l-i} + 2\alpha^2 \gamma^{l-i}$

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 $\mathbb{E}[\deg(u)|u \in W_t]$ 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 $\rho$ might be validated by the next two corollaries.

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

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

Note that $\log(\mu^2 \alpha + 2 \mu (1-\mu) \beta + (1-\mu)^2 \gamma) < 0$ because both $\mu$ and $\gamma$ 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 following way. When a new node joins the network, its behavior is governed by the node attribute distribution, which is seemingly independent of the graph structure. However, in the long term, since the number of attributes grows slowly as the number of nodes increases, the node attributes and the graph 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 constitute 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 $\frac{1}{2}n^2 + \rho \log(n^2 \alpha + 2 \mu (1-\mu) \beta + (1-\mu)^2 \gamma)$. We can easily figure out that this fact agrees with the Densification Power Law [23], 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 desenification exponent $a = \log(|\Theta|)$ where $|\Theta|$ denotes the sum of all entries in $\Theta$.

The full proofs of all are 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 $\left( (\mu^2 \alpha + (1-\mu) \beta)^{\mu} (\mu \beta + (1-\mu) \gamma)^{1-\mu} \right)^{\rho} \geq \frac{1}{2}$

**Theorem 4.2** (Connectedness) $M(n, l, \mu, \Theta)$ is connected with high probability as $n \to \infty$, if

$$F_C(M) = \begin{cases} 
(\mu \beta + (1-\mu) \gamma)^{\rho} & \text{when } (1-\mu)^\rho \geq \frac{1}{2} \\
(\mu^2 \alpha + (1-\mu) \beta)^{\nu} (\mu \beta + (1-\mu) \gamma)^{1-\nu} & \text{otherwise}
\end{cases}$$

is greater than $\frac{1}{2}$ where $0 < \nu < \mu$ is a solution of the equation $\left( \left( \frac{\mu}{2} \right)^{\nu} \left( \frac{1-\mu}{2} \right)^{1-\nu} \right)^{\rho} = \frac{1}{2}$.

In contrast, $M(n, l, \mu, \Theta)$ is disconnected with high probability as $n \to \infty$ if $F_C(M) < \frac{1}{2}$. 

7
To show the above theorems, we first define the monotonicity property of MAG model.

**Theorem 4.3 (Monotonicity)** For \( u, v \in V \), \( \mathbb{E} [ P[u,v] | ||u|| = i] \leq \mathbb{E} [ P[u,v] | ||u|| = j] \) if \( i \leq j \)

Theorem 4.3 ultimately demonstrates that the node of larger weight is more likely to be connected with other nodes. In another view, the 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 connectivity as well as on the existence of giant component.

By the monotonicity property, the minimum degree is likely to be the degree of the minimum weight node. Therefore, the disconnectivity 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 criteria in Theorem 4.2 are separated into two cases under the condition on \( \mu \), which tells whether or not the expected number of weight 0 nodes, i.e., \( \mathbb{E} [|W_0|] \), 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 \( \nu \) describes the ratio of the minimum weight to \( l \) as \( n \to \infty \). Therefore, the condition for connectedness is actually dependent 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 by using some algebra introduced in [26].

Similar explanation may work for the existence of a giant component. Instead of the minimum weight node, Theorem 4.1 describes that the existence of \( \Theta(n) \) component relies on the degree of the median weight node. We can 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 3 different weights: \( \mu l \), \( \mu l + l^{1/6} \), 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 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 [26].

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

This theorem does not specify the exact value of diameter, but, at least under the given condition, it guarantees the bounded diameter even though \(n \to \infty\). The proof is based on the following theorem.

**Theorem 5.2** [7] [16] 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 \(n \to \infty\).
Theorem 5.2 describes only Erdős-Rényi random graph. However, if we can assure that the edge probability between any pair of nodes in a MAG graph is greater than $p$, the diameter of that graph would be at most that of $G(n, p)$ [26]. Using this feature combined with Theorem 5.2, we can derive main lemmas as follow:

**Lemma 5.3** 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 5.4** If $(\mu \beta + (1 - \mu) \gamma)^\rho > \frac{1}{2}$, for $\lambda = \frac{\mu \beta}{\mu \beta + (1 - \mu) \gamma}$, all nodes in $V \setminus S_{\lambda l}$ are directly connected to $S_{\lambda l}$ with high probability as $n \to \infty$.

By Lemma 5.4, we can conclude that the diameter of the entire graph is limited to $(2 + \text{diameter of } S_{\lambda l})$.

Since by Lemma 5.3 the diameter of $S_{\lambda l}$ is constant with high probability under the given condition, the actual diameter is also constant.

The detailed 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. Depending on $\Theta$, MAG model produces graphs of various degree distributions. For instance, since the network becomes a sparse Erdős-Rényi random graph if $\alpha \approx \beta \approx \gamma < 1$, the degree distribution will approximately follow the binomial distribution. For another extreme example, in case of $\alpha \approx 1$ and $\mu \approx 1$, the network will be close to a clique, which represents a degree distribution different from a sparse Erdős-Rényi random graph. For this reason, we need to narrow down the conditions on $\mu$ and $\Theta$ as follows.

If $\mu$ is close to 0 or 1, then the graph becomes an Erdős-Rényi random graph with edge probability $p = \alpha$ (when $\mu \approx 1$) or $\gamma$ (when $\mu \approx 0$). Since the degree distribution of Erdős-Rényi random graph is binomial, we will exclude these extreme cases of $\mu$. On the other hand, with regard to $\Theta$, We assume that a reasonable configuration space for $\Theta$ 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 $\Theta$ can be therefore supported by real examples in [22]. Assuming all these conditions, we reach the following theorem about the degree distribution.

**Theorem 6.1** Under above assumptions, if $\left[(\mu \alpha + (1 - \mu) \beta) \mu \beta + (1 - \mu) \gamma\right]^{\rho} > \frac{1}{2}$, then the tail of degree distribution $p_k$ of $M(n, l, \mu, \Theta)$ follows a log-normal as $n \to \infty$, for $R = \frac{\mu (1 - \mu) \alpha}{\mu \beta + (1 - \mu) \gamma}$

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

In other words, the degree distribution of MAG model approximately illustrates the quadratic relationship on log-log scale. This result is acceptable since some social networks follow the log-normal distribution. For instance, the degree distribution of LiveJournal network looks more parabolic than linear on this scale [23].

In brief, since the expected degree is an exponential function of the node weight by Lemma 5.3, the degree distribution might be mainly affected by the distribution of node weight. Whereas this node weight

---

<sup>3</sup> We trivially exclude self-edges not only because computations become simple but also because other models usually do not include them.
actually follows binomial, it can be approximated by 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 normal distribution for large \( l \). This eventually indicates that the degree distribution roughly follows a log-normal.

Note that there exists a condition, \( \left[ (\mu \alpha + (1 - \mu) \beta)^{\mu} \left( \mu \beta + (1 - \mu) \gamma \right)^{1 - \mu} \right]^{\theta} > \frac{1}{2} \), 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 represents that the median degree is large enough. Equivalently, it also indicates that the degrees of a half of 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 show Theorem 6.1.

To analyze it more rigorously, we need the following theorem and its corollary.

**Theorem 6.2** \[37\] \( P \left( \text{deg}(u) = k \right) = \int_{u \in V} \left( \binom{n-1}{k} \left( \mathbb{E} \left[ P \left[ u, v \right] \right] \right)^k \left( 1 - \mathbb{E} \left[ P \left[ u, v \right] \right] \right)^{n-1-k} \right) \, du \)

**Corollary 6.2.1** 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 \( p_k = \sum_{j=0}^{l} \binom{l}{j} (\mu)^{j} (1 - \mu)^{l - j} \binom{n-1}{k} E_j^k (1 - E_j)^{n-1-k} \)

Corollary 6.2.1 can be easily obtained by combining Lemma 5.2 with Theorem 6.2, but it seems difficult to find the exact closed form expression. Therefore, we need some approximations such as Stirling approximation and normal approximation to prove Theorem 6.1. In more detail, we seek the dominant term in Corollary 6.2.1, and then show that this term governs \( p_k \) and the log value of it is approximately a quadratic function of \( \ln k \). The quadratic function of \( \ln k \) eventually represents the log-normality.

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 \[13\].

In this section, we show that 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 do not equalize the attribute distributions as well as the similarity matrices. The formal definition of this model is as follows:

\[
P(a_j(u) = 1) = \mu_j, \quad P(u, v) = \prod_{j=1}^{l} \Theta_j[a_j(u), a_j(v)]
\]

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

**Theorem 7.1** For \( M(n, l, \tilde{\mu}, \tilde{\Theta}) \), 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 there exists some configurations of \( \tilde{\mu} \) and \( \tilde{\Theta} \) that 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 a attribute vector $a(u)$ is

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

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

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

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.

To prove this theorem in detail, we can apply similar methods to Section 6. First, we compute the exact expression of $p_k$ based on Theorem 6.2. Next, we approximate $p_k$ to the dominant term by using some algebra. Finally, we obtain the expression of this dominant term, which is roughly proportional to $k^{-\delta-\frac{1}{2}}$.

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

### 8 Simulation

In previous sections, we performed theoretical analyses of the MAG model. In this section, we use simulation experiments to further demonstrate the properties of networks that arise under the MAG graph. First, we generated the 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 fraction of nodes in the largest connected component of the graph. Second, we also ran simulations with fixed parameter configurations to check the other network properties. In this way, we are able to qualitatively compare our model to a real-world network.

**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 $\mu, \alpha, f$, and $n$, where $f$ indicates the scalar factor of $\Theta$, i.e., $\Theta = f \cdot \Theta_0$ for a constant $\Theta_0 = [\alpha \beta; \beta \gamma]$.

Figure 2 depicts the number of edges, the fraction of nodes in the largest connected component, and the effective diameter of the network as a function $\mu, \alpha, f$, 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 the figure with log scaled $y$-axis. Second, 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, like in the Erdős-Rényi random graph model [12]. Last, while the previous two network properties monotonically change, 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 about the evolution of real-world networks [23, 27] observed the “gelling” point where the giant component forms and the diameter starts to decrease.
the local clustering of nodes would naturally emerge by mixing core-periphery matrices \((\alpha > \beta > \gamma)\) for all attributes. Thus, all attributes model the overall core-periphery shape of the Yahoo!-Flickr network, while the local clustering is the effect of the opposite process of homophily and network community formation. Our hypothesis is that the local clustering of nodes would naturally emerge by mixing core-periphery \(\Theta_i\) matrices \((\alpha > \beta > \gamma)\).

**Figure 2:** Structural properties of a simplified MAG graph when we vary a single parameter: \(\mu, \alpha, f\), or \(n\). As the parameter increases, the graph becomes denser so that a giant component emerges, whereas the diameter decreases to approach a constant.

Furthermore, we also performed simulations where we keep \(\Theta\) and \(\mu\) constant but simultaneously increasing both \(n\) and \(l\) by maintaining their ratio fixed. Figure 3 plots the change in each metric as a function of \(n\) for different values of \(\mu\), and effectively represents the evolution of the MAG network over time. From the plots, we see that MAG model follows densification power law (DPL) and the shrinking diameter.

**Comparison to Real-world Networks.** Also, we qualitatively compare the structural properties of a real-world network and the corresponding MAG model. For these experiments, we simply selected some parameter settings (for \(n, l, \mu, \Theta\)) 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 here is not to claim that these particular parameter values are in any way “optimal” for the given real-world network but to rather show that many properties of the MAG model exhibit qualitatively similar behavior as real-world networks.

For this comparison, we use a real-world Yahoo!-Flickr online social network on 10,000 nodes and 44,800 edges. Figure 4 illustrates various properties of the real and the corresponding synthetic network and reveals that the plots of MAG model resemble those of Yahoo!-Flickr network. Notice qualitatively similar behavior of nearly all properties. The only property where the MAG network does not match the Yahoo!-Flickr network seems to be the clustering coefficient. As in real-world networks high degree nodes tend to have higher clustering, in the simplified MAG model the situation is reversed – higher degree nodes also tend to have higher clustering. This is due to the fact that we are using same \(\Theta\) for all attributes. Thus, all attributes model the overall core-periphery shape of the Yahoo!-Flickr network, while the local clustering is the effect of the opposite process of homophily and network community formation. Our hypothesis is that the local clustering of nodes would naturally emerge by mixing core-periphery \(\Theta_i\) matrices \((\alpha > \beta > \gamma)\).
and homophily $\Theta_i$ matrices ($\alpha, \gamma > \beta$).

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.
Figure 4: The comparison of network properties between real-world Yahoo!-Flickr online social network and a simplified MAG model network. Except for clustering coefficient the properties of MAG graph qualitatively resemble those of Yahoo!-Flickr network.

[9] 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.

[10] 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.

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

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

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

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

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

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

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

[18] 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.
Figure 5: Degree distributions of simplified and power-law version of MAG graphs. We plot both PDF and CCDF of the degree distribution, while the simplified version (a) has parabolic shape log-log scale, which is an indication of a log-normal distribution, while figure (b) demonstrates a power-law relationship.

[19] S. Lattanzi and D. Sivakumar. Affiliation networks. In STOC ’09, pages 427–434, 2009.

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

[21] 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.

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

[23] 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.

[24] 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.

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

[26] M. Mahdian and Y. Xu. Stochastic kronecker graphs. In WAW ’07, 2007.

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

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

[29] 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.

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

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

[32] M. Mitzenmacher. Editorial: The future of power law research. Internet Mathematics, 2:525–534, 2006.
[33] G. Palla, L. Lovász, and T. Vicsek. Multifractal network generator. PNAS, 107(17):7640–7645, 2010.

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

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

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

[37] 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 attribute 0’s shared by \( u \) and \( v \). In a similar way, we can define \( N_{uv}^1 \) as the number of attribute 1’s shared. Clearly, \( N_{uv}^0, N_{uv}^1 \geq 0 \) and \( N_{uv}^0 + N_{uv}^1 \leq l \) hold. Then, by the definition of MAG model, the edge probability between \( u \) and \( v \) is \( \alpha N_{uv}^1 \beta l N_{uv}^0 \).

Since \( N_{uv}^i \)'s are random variables, in order to compute the edge probability, we need some kind of their joint distribution. Here we find the conditional joint distribution of \( N_{uv}^i \)'s given the weight of \( u \). Because \( N_{uv}^0 \) and \( N_{uv}^1 \) are independent each other, \( N_{uv}^0 \) eventually represents the number of heads in \( (l-i) \) coin tosses with probability \((1 - \mu)\), which follows Bin\((l - i, 1 - \mu)\). Similarly, \( N_{uv}^1 \) follows Bin\((i, \mu)\). Therefore, their conditional joint probability is

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

Using this, we can compute the expectation of \( P[u, v] \) given the weight of \( u \):

\[
\mathbb{E}[P[u, v] | u \in W_i] = \mathbb{E} \left[ \alpha^{N_{uv}^1} \beta^{l - N_{uv}^1} \beta^{-i - N_{uv}^0} \gamma^{N_{uv}^0} | u \in W_i \right]
\]

\[
= \sum_{N_{uv}^0 = 0}^{l-i} \sum_{N_{uv}^1 = 0}^{i} \binom{i}{N_{uv}^1} \mu^{N_{uv}^1} (1 - \mu)^{i - N_{uv}^1} \binom{l - i}{N_{uv}^0} \mu^{l - i - N_{uv}^0} (1 - \mu)^{N_{uv}^0}
\]

\[
= (\mu \alpha + (1 - \mu) \beta)^i (\mu \beta + (1 - \mu) \gamma)^{l - i - i}
\]

Proof of Lemma 3.3: By Lemma 3.2 and the linearity of expectation, we can sum this conditional probability over all nodes to result in the conditional expectation of the degree.

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

\[
\mathbb{E}[m] = \mathbb{E} \left[ \frac{1}{2} \sum_{u \in V} \text{deg}(u) \right]
\]

\[
= \frac{1}{2} \sum_{j=0}^{l} P(W_j) \mathbb{E}[\text{deg}(u)|u \in W_j]
\]

\[
= \frac{1}{2} n \sum_{j=0}^{l} \binom{l}{j} \mu^j (1 - \mu)^{l - j} \mathbb{E}[\text{deg}(u)|u \in W_j]
\]

\[
= \frac{1}{2} n \sum_{j=0}^{l} \binom{l}{j} \left( (n - 1)(\mu \alpha + (1 - \mu) \beta)^j (\mu \beta + (1 - \mu) \gamma)^{l - j} + 2 \alpha \mu j \gamma^{l - j} (1 - \mu)^{l - j} \right)
\]

\[
= \frac{n(n - 1)}{2} \left( \mu^2 \alpha + 2 \mu (1 - \mu) \beta + (1 - \mu)^2 \gamma \right) + n (\mu \alpha + (1 - \mu) \gamma)
\]
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 \( \Theta(n^2 \zeta^l) \). Since \( \zeta < 1 \) and \( \log \zeta < 0 \), this expectation is after all \( \Theta(n^2 \zeta^l) = \Theta(\zeta^l + 2 \log n \log \zeta) = \Theta(n^{1+\epsilon \log \zeta}) = o(n) \).

Proof of Corollary 3.3.2: Under the situation that \( l \in o(\log n) \), the expected degree is \( \Theta(n^2 \zeta^l) = \Theta(n^{2+\log n \log \zeta}) = \Theta(n^{2-o(1)}) \).

B Appendix: Connectivity

Proof of Theorem 4.3: Since \( j \geq i \), for any \( v_i \in W_i \), we can easily find a node \( v_i^{(j)} \) by flipping \( j - i \) zero bits randomly in \( v_i \) so that \( P[u, v_i^{(j)}] \geq P[u, v_i] \). Then, it is obvious that \( \mathbb{E}[P[u, v_i^{(j)}] | v_i] \geq \mathbb{E}[P[u, v_i]] \).

Therefore,

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

Lemma B.1 \( V_{\min} \to 0 \) with high probability as \( n \to \infty \) if \( (1 - \mu)^\rho \geq \frac{1}{2} \). Otherwise, \( V_{\min} \to \nu \) with high probability as \( n \to \infty \) where \( \nu \) is a solution of the equation \( \left( \frac{\nu}{\mu} \right)^\rho \left( \frac{1}{1-\mu} \right) = \frac{1}{2} \) and \( \nu < \mu \).

Proof: First, we assume that \( (1 - \mu)^\rho \geq \frac{1}{2} \). The probability that \( |W_i| = 0 \) is at most \( \exp(-\frac{1}{2} \mathbb{E}[|W_i|]) \) by Chernoff bound. Since

\[
\mathbb{E}[|W_i|] = n \binom{l}{1} \mu^l (1 - \mu)^{l-1} \geq \frac{\mu l}{1 - \mu}
\]

, \( P(|W_1| = 0) \) goes to zero as \( l \to \infty \). Therefore, \( V_{\min} \) is \( O(1) \) with high probability.
Theorem 4.3

a subset with high probability if between the weight of which is less than

\[ (1 - \mu)^\rho < \frac{1}{2} \]

For any \( \mu - \nu > \epsilon > 0 \),

\[
\mathbb{E} \left[ |W_{(\nu+\epsilon)}| \right] \approx n \left( \frac{l}{(\nu + \epsilon)} \right)^{(\nu+\epsilon)l} (1 - \mu)^{(1-(\nu+\epsilon))l} \]

\[
\approx \frac{n \mu^{(\nu+\epsilon)l}(1 - \mu)^{(1-(\nu+\epsilon))l} \sqrt{2\pi l (\nu + \epsilon)}^{(\nu+\epsilon)l} \sqrt{2\pi l (1 - (\nu + \epsilon))^{(1-(\nu+\epsilon))l}}}{\sqrt{2\pi l (\nu + \epsilon)(1 - (\nu + \epsilon))^{(1-(\nu+\epsilon))l}}} \]

Since \( \left( \frac{l}{n} \right)^x \left( \frac{1}{1-x} \right)^{-x} \) is a increasing function of \( x \) over \( (0, \mu) \),

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

for a constant \( \epsilon' > 0 \). Therefore,

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

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

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

To sum up, as \( n \) and \( l \) increases, \( V_{\min} \) tends to be \( \nu l \) with high probability as \( l \to \infty \).

Proof of Theorem 4.2: Assume that \( |S_j| \in \Theta(n) \) for some \( j \). Then, we hope to claim that \( S_j \) is connected with high probability if \( \mathbb{E} \left[ P \left[ u, V \setminus u \right] \right] \geq c \log n \) for sufficiently large \( c \) as \( n \to \infty \). Let’s think of a subset \( S' \subset S_j \) such that \( S' \) is neither an empty set nor \( S_j \) itself. Then, the expected number of edges between \( S' \) and \( S_j \setminus S' \) is \( \mathbb{E} \left[ P \left[ S', S_j \setminus S' \right] \right] = |S'| \cdot |S_j - S'| \cdot \mathbb{E} \left[ P \left[ u, v \right] \right] \) for distinct \( u \) and \( v \). By Theorem 4.3,

\[
\mathbb{E} \left[ P \left[ u, v \right] \right] \geq \mathbb{E} \left[ P \left[ u, v \right] \right] \geq \mathbb{E} \left[ P \left[ u, v \right] \right] \geq \frac{c \log n}{n}
\]

Since the probability that there exists no edge between \( S' \) and \( S_j \setminus S' \) is at most \( \exp \left( - \frac{1}{2} \mathbb{E} \left[ P \left[ S', S_j \setminus S' \right] \right] \right) \)



19
by Chernoff bound, it is bounded as follows:

\[
P(S \text{ is disconnected}) \leq \sum_{S' \subset S, S' \neq S_j} P(\text{no edge between } S', S_j \setminus S')
\]

\[
\leq \sum_{S' \subset S, S' \neq S_j} \exp\left(-\frac{c \log n}{2n} |S'| \cdot |S_j \setminus S'| \right)
\]

\[
\leq 2 \sum_{1 \leq i \leq |S_j|/2} \exp\left(\frac{|S_j|}{i} \exp\left(-\frac{c |S_j| \log n}{2n} i \right)\right)
\]

\[
= 2 \sum_{1 \leq i \leq |S_j|/2} \exp\left(-O(i \log n)\right) \in o(1)
\]

as \( n \to \infty \). Therefore, \( S_j \) is connected with high probability.

Let \( \frac{V_{\min}}{\log n} \to t \) for a constant \( 0 \leq t < \mu \) as \( n \to \infty \). Then, this setting can cover both conditions by Lemma B.1.

If \( (\mu \alpha + (1 - \mu) \beta)^t (\mu \beta + (1 - \mu) \gamma)^{1-t} \rho > \frac{1}{2} \), then

\[
\mathbb{E}[P[u, V \setminus u] | u \in W_{V_{\min}}] \approx \mathbb{E}[P[u, V \setminus u] | u \in W_t]
\]

\[
\approx \left[2 \left[(\mu \alpha + (1 - \mu) \beta)^t (\mu \beta + (1 - \mu) \gamma)^{1-t} \right] \right]^{\log n}
\]

is greater than \( c \log n \). Since \( |S_{V_{\min}}| \) is clearly \( \Theta(n) \), \( S_{V_{\min}} \) is connected with high probability by the proceeding arguments. By the definition of \( V_{\min} \), the entire graph is also connected with high probability.

On the other hand, when \( (\mu \alpha + (1 - \mu) \beta)^{V_{\min}/\log n} (\mu \beta + (1 - \mu) \gamma)^{\frac{l - V_{\min}}{\log n}} < \frac{1}{2} \), the expected degree of a node with \( |V_{\min}| \) weight is \( o(1) \) from the above relationship. Thus, some node in \( W_{V_{\min}} \) is isolated with high probability in this case.

\section{Appendix: Giant Component}

The following lemmas tell us the size of each subgraph.

\begin{lemma}
\label{lem:superior}
\|S_{\mu l} \| \geq \frac{n}{2} - o(n) \text{ as } n \to \infty.
\end{lemma}

\textbf{Proof:} By Central Limit Theorem, \( |u| - \mu l \sim \sqrt{l \mu(1 - \mu)} N(0, 1) \) as \( l \to \infty \). Therefore, \( P(|u| \geq \mu l) \) is at least \( \frac{1}{2} - o(1) \). Then, by the Law of Large Number, \( |S_{\mu l}| \) is at least \( \frac{n}{2} - o(n) \) with high probability as \( l \to \infty \), i.e., \( n \to \infty \).

\begin{lemma}
\label{lem:superior1}
\|S_{\mu l + l^{1/6}} \| \in \Theta(n) \text{ as } n \to \infty.
\end{lemma}

\textbf{Proof:} By Central Limit Theorem mentioned in Lemma C.1

\[
P(\mu l \leq |u| < \mu l + l^{1/6}) \approx \Phi\left(\frac{\mu l + l^{1/6}}{\sqrt{l \mu(1 - \mu)}}\right) - \Phi(0)
\]

20
is \( o(1) \) 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 \( \Theta(n) \) as \( l \to \infty \), i.e., \( n \to \infty \).

\[ \text{Lemma C.3} \quad |S_{\mu l + l^{2/3}}| \in o(n) \text{ as } n \to \infty. \]

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

**Proof of Theorem 4.1:** (Existence) First, if \[ \left( (\mu \alpha + (1 - \mu) \beta)^{\mu} (\mu \beta + (1 - \mu) \gamma)^{1-\mu} \right)^{(n \log n)^{1/6}} > \frac{1}{2}, \]
then
\[ \mathbb{E} \left[ P[u, V \backslash u] \mid u \in W_{\mu l}\right] > c \log n \]
holds. Since \( |S_{\mu l}| \in \Theta(n) \) by Lemma C.3, \( S_{\mu l} \) is connected with high probability by Theorem 4.2. In other words, we are able to extract out a connected component of size at least \( \frac{n}{2} - o(n) \).

Second, when \[ \left( (\mu \alpha + (1 - \mu) \beta)^{\mu} (\mu \beta + (1 - \mu) \gamma)^{1-\mu} \right)^{(n \log n)^{2/3}} = \frac{1}{2}, \]
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 C.2 and
\[ \mathbb{E} \left[ P[u, V \backslash u] \mid u \in W_{\mu l + l^{1/6}} \right] \approx \left( \frac{\mu \alpha + (1 - \mu) \beta}{\mu \beta + (1 - \mu) \gamma} \right)^{(n \log n)^{1/6}} \]
is also greater than \( c \log n \). Thus, \( S_{\mu l + l^{1/6}} \) is connected with high probability by Theorem 4.2.

Last, on the contrary, when \[ \left( (\mu \alpha + (1 - \mu) \beta)^{\mu} (\mu \beta + (1 - \mu) \gamma)^{1-\mu} \right)^{(n \log n)^{2/3}} < \frac{1}{2}, \]
for \( u \in W_{\mu l + l^{2/3}} \),
\[ \mathbb{E} [P[u, V \backslash u]] \approx \left( \left( (\mu \alpha + (1 - \mu) \beta)^{\mu} (\mu \beta + (1 - \mu) \gamma)^{1-\mu} \right)^{(n \log n)^{2/3}} \log n \right) \left( \frac{\mu \alpha + (1 - \mu) \beta}{\mu \beta + (1 - \mu) \gamma} \right)^{(n \log n)^{2/3}} \]
is \( o(1) \) as \( n \to \infty \). Since \( S_{\mu l + l^{2/3}} \) is \( o(n) \) by Lemma C.3, most of \( n - o(n) \) nodes are isolated, therefore the size of the largest component cannot be \( \Theta(n) \).

(Uniqueness) We already pointed out that \( S_{\mu l} \) or \( S_{\mu l + l^{1/6}} \) can be the \( \Theta(n) \) component under the certain condition. Let this component be \( H \). Then, since \( \mathbb{E} [P[u, v] \mid v \in H] \geq \mathbb{E} [P[u, v] \mid v \in V \backslash H] \) by Theorem 4.3,
\[ \mathbb{E} [P[u, V \backslash H]] \leq \left( \frac{n - |H|}{|H|} \right) \mathbb{E} [P[u, H]] \]
holds for every \( u \in V \).

Assume that another connected component \( H' \) also contains \( \Theta(n) \) nodes. As described before, we will show that it is connected to \( H \) with high probability by presenting the contradiction.
\[ \mathbb{E} \left[ P[H, H'] \right] = |H'| \cdot \mathbb{E} \left[ P[u, H] \mid u \in H' \right] \geq \frac{|H'|}{n - |H|} \mathbb{E} \left[ P[u, V \backslash H] \mid u \in H' \right] \geq \frac{|H'|}{n - |H|} \mathbb{E} \left[ P[u, H] \mid u \in H' \right] \]
Since both \( |H| \) and \( |H'| \) are \( \Theta(n) \), if \( \mathbb{E} [P[u, H'] \mid u \in H'] \) is \( \Omega(1) \), then \( \mathbb{E} [P[H, H']] \in \Omega(n) \), which indicates that \( H \) and \( H' \) is connected with high probability. On the other hand, if \( \mathbb{E} [P[u, H'] \mid u \in H'] \in \mathcal{O}(1) \), then \( H' \) should have at least one isolated node with high probability by Chernoff bound. This is also contradiction. To sum up, there is no more \( \Theta(n) \) connected component with high probability.
D Appendix: Diameter

Proof of Lemma 5.4: Since \( \min_{u,v \in S_{\lambda l}} P[u,v] \geq \beta^{\lambda l}(1-\lambda)^l \), it is enough to prove that \( G(|S_{\lambda l}|, \beta^{\lambda l}(1-\lambda)^l) \) has a constant diameter. However,

\[
E [|W_{\lambda l}|] = n \left( \begin{array}{c} l \\ \lambda l \end{array} \right) \mu^{\lambda l}(1-\mu)^{(1-\lambda)^l} \beta^{\lambda l}(1-\lambda)^l \\
\approx \frac{n}{\sqrt{2\pi l \lambda (1-\lambda)}} \left( \frac{\mu \beta}{\lambda} \right)^{\lambda l} \left( \frac{(1-\mu) \gamma}{1-\lambda} \right)^{(1-\lambda)^l} \\
= \frac{n}{\sqrt{2\pi l \lambda (1-\lambda)}} (\mu \beta + (1-\mu) \gamma)^l \\
= \frac{1}{\sqrt{2\pi l \lambda (1-\lambda)}} (2 (\mu \beta + (1-\mu) \gamma)^l)^{\log n}
\]

Since \( |W_{\lambda l}| \) is \( \Theta(E[W_{\lambda l}]) \) as \( n \to \infty \), \( |S_{\lambda l}| (\min_{u,v \in S_{\lambda l}} P[u,v]) \) is at least \( \Theta \left( \frac{(1+\epsilon \log n}{\sqrt{l}} \right) \) for a constant \( \epsilon > 0 \).

By Theorem 5.2, an Erdős-Rényi random graph \( G(|S_{\lambda l}|, \frac{(1+\epsilon \log n}{|S_{\lambda l}|^{1/l}} \) has a diameter of at most \( (1 + \frac{1}{\epsilon}) \) as \( n \to \infty \). Thus, the diameter of \( S_{\lambda l} \) is also bounded by a constant.

Proof of Lemma 5.4: For any \( u \in V \),

\[
E[P[u, S_{\lambda l}]] = \sum_{j=\lambda l}^{l} n \left( \begin{array}{c} l \\ j \end{array} \right) \mu^{j}(1-\mu)^{l-j} \beta^{j}(1-\lambda)^{l-j} \\
= (2 (\mu \beta + (1-\mu) \gamma)^l)^{\log n} \left( \sum_{j=\lambda l}^{l} \left( \begin{array}{c} l \\ j \end{array} \right) \lambda^{j}(1-\lambda)^{l-j} \right)
\]

By Central Limit Theorem, \( \sum_{j=\lambda l}^{l} \left( \begin{array}{c} l \\ j \end{array} \right) \lambda^{j}(1-\lambda)^{l-j} \) converges to \( \frac{1}{2} \) as \( l \to \infty \). Therefore, \( E[P[u, S_{\lambda l}]] \) 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.

E Appendix: Degree Distribution

Proof of Theorem 6.1: To reduce the space, we first define useful notations as follow:

\[
x = \mu \alpha + (1-\mu) \beta \\
y = \mu \beta + (1-\mu) \gamma \\
f_j(k) = \left( \begin{array}{c} n-1 \\ k \end{array} \right) (x^j y^{l-j})^k \left( 1-x^j y^{l-j} \right)^{n-1-k} \\
g_j(k) = \left( \begin{array}{c} l \\ j \end{array} \right) \mu^{j}(1-\mu)^{l-j} f_j(k)
\]

By Corollary 6.2.1 we can restate \( p_k \) as \( p_k = \sum_{j=0}^{l} g_j(k) \).
If most of those terms turn out to be insignificant under our assumptions, the probability \( p_k = P(deg(u) = k) \) could be approximately proportional to one or few dominant terms. In this case, what we need to do is thus to seek for \( j \) such that maximizes \( g_j(k) = \binom{\ell}{j} \mu^j (1 - \mu)^{\ell - j} f_j(k) \) and find its approximate formula.

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

\[
f_j(k) \approx \frac{\sqrt{2\pi n}(n/e)^n (x^j y^{l-j})^k (1 - x^j y^{l-j})^{n-k}}{\sqrt{2\pi k(n/e)^k} \sqrt{2\pi(n-k)((n-k)/e)^{n-k}}} = \frac{1}{2\pi k \left(1 - \frac{k}{n}\right)} \left( \frac{nx^j y^{l-j}}{k} \right)^k \left( 1 - \frac{x^j y^{l-j}}{1-k/n} \right)^{n-k}
\]

However, because the degree of maximum weight node is expected to be \( O(n (\mu\alpha + (1 - \mu)\beta)^{\ell}) \), \( k \) would be \( o(n) \) with high probability.

\[
: \quad \left(1 - \frac{x^j y^{l-j}}{1-k/n}\right)^{n-k} \approx \exp(-(n-k)x^j y^{l-j} + (n-k)k/n) \approx \exp(-nx^j y^{l-j} + k)
\]

For large \( l \), we can further solve \( g_j(k) \) by normal approximation of the binomial distribution:

\[
\ln g_j(k) \approx C - \frac{1}{2} \frac{(\mu(1-\mu)(j - \mu)^2 - \frac{1}{2}\ln k - k \ln \frac{k}{nx^j y^{l-j}} + k \left(1 - \frac{x^j y^{l-j}}{k}\right)}{2\mu(1-\mu)} = C - \frac{1}{2} \frac{(j - \mu)^2}{2\mu(1-\mu)} - \frac{1}{2} \ln k + k(j - \tau) \ln R + k \left(1 - R^{l-\tau}\right)
\]

for \( k = nx^\tau y^{l-\tau} \) (\( \mu \leq \tau \)) and a constant \( C \). Using \((j - \mu)^2 = (j - \tau)^2 + (\tau - \mu)^2 + 2(j - \tau)(\tau - \mu)\),

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

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

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

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

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

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

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

Furthermore, since \(|\frac{j - \tau}{l\mu(1-\mu)}| \ll k \ln R \) as \( n \to \infty \), the first quadratic term in \( \ln g_j(k) \) is negligible. In the 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 \( \Delta \geq 1 \).

23
Thus, what we try to show is that for a node $\kappa$, i.e., $n p_k$ is roughly proportional to $n g_r$. By assigning $\tau = \frac{\ln k - \ln n y^l}{\ln R}$, we are able to obtain

$$
\ln p_k \approx C - \frac{1}{2l\mu(1-\mu)} \left( \frac{\ln k - \ln n y^l}{\ln R} - \mu \right)^2 - \frac{1}{2} \ln k
$$

$$
= C' - \frac{1}{2l\mu(1-\mu)(\ln R)^2} \left( \ln k - \ln n y^l - l\mu \ln R - \frac{1}{2} l\mu(1-\mu)(\ln R)^2 \right)^2 - \ln k
$$

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

\section*{Appendix: Power-law Distribution}

\textbf{Proof of Lemma 7.2}: Since $a_i$’s are independently distributed Bernoulli random variables, Lemma 7.2 holds.

\textbf{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.,

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

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

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

First, when $l = 1$, it is trivially true by Lemma 3.2.

Second, when $l > 1$, since $P_l(u, v) = P_{l-1}(u, v) \Theta_l[a_l(u), a_l(v)]$, we show that if this holds when $l = k - 1 \geq 1$ then it also holds when $l = k$. For a node $v \in V \setminus u$,

$$
\mathbb{E} [P_k(u, v)] = \mathbb{E} [P_{k-1}(u, v)] \mathbb{E} [\Theta_k[a_k(u), a_k(v)]]
$$

$$
= \mathbb{E} [P_{k-1}(u, v)] (\mu_k \alpha_k + (1 - \mu_k) \beta_k)^{1\{a_k(u)=1\}} (\mu_k \beta_k + (1 - \mu_k) \gamma_k)^{1\{a_k(u)=0\}}
$$

$$
= \prod_{i=1}^{k} (\mu_i \alpha_i + (1 - \mu_i) \beta_i)^{1\{a_i(u)=1\}} (\mu_i \beta_i + (1 - \mu_i) \gamma_i)^{1\{a_i(u)=0\}}
$$

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

\textbf{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 $p(1) = 0.1$, $p(2) = 0.2$, and so on.

Then, by Theorem 6.2, we can express the probability of degree $k$, $p_k$, as follows:

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

(2)
where $E_j$ denotes the expected edge probability of attributes 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 $\binom{n-1}{k} p(\tau) E^{k}(1-E)^{n-1-k}$ 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.

For the simplicity of computations, we propose a condition that $p(j) \propto E_{j}^{-\delta}$ for a constant $\delta$. Since the $j$-th term is proportional to $(E_j)^{k-\delta} (1-E_j)^{n-1-k}$, it 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 $n \rightarrow \infty$. Therefore, the $\tau$-th term dominates the Equation (2).

Next, by the Stirling approximation with above conditions, $p_k \propto (1 - \frac{\delta}{k})^k \left(1 - \frac{\delta}{n-1}\right)^{n-1-k}$ holds for large $k$ and $n$. Thus, $\ln p_k$ is proportional to $k^{-\frac{1}{2} - \delta}$ for large $k$ as $n \rightarrow \infty$.

Last, we prove that the two conditions for the power-law degree distribution are simultaneously feasible by providing an example setting. If every $p(j)$ is distinct and $\frac{\mu_i}{1-\mu_i} \propto \left(\frac{\mu_i, \alpha_i + (1-\mu_i) \beta_i}{\mu_i, \beta_i + (1-\mu_i) \gamma_i}\right)^{-\delta}$, then we are able to satisfy the condition that $p(j) \propto (E_j)^{-\delta}$ by Lemma [7.2] and Lemma [7.3]. On the other hand, if we can set $\frac{\mu_i}{1-\mu_i} = (C_z (1 + z))^{-2i \delta}$ with a proper constant $C_z$, then the other condition, $E_{j+1}/E_{j} \geq (1 + z)$ is also satisfied. Since we are free to configure $\mu_i$’s and $\Theta_i$’s independently, the sufficient condition for the power-law degree distribution is feasible.