LLNL-SR-460642

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Multiplicative Attribute Graph Model of Real-World Networks

M. Kim, J. Leskovec

October 22, 2010

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

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, is ubiquitous in a variety of scientific domains. The study of such social and information networks commonly finds patterns and explain their emergence through tractable models. In most networks, especially in social networks, nodes also have a rich set of attributes (*e.g.*, age, gender) associated with them. However, most of the existing network models focus only on modeling the network structure while ignoring the features of nodes in the network.

Here we present a class of network models 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 features associated with it. The probability of an edge between a pair of nodes then depends on the product of individual attribute-attribute similarities. The model yields itself to mathematical analysis as well as fit to real data. We derive thresholds for the connectivity, the emergence of the giant connected component, and show that the model gives rise to graphs with a constant diameter. Moreover, we analyze the degree distribution to show that the model can produce networks with either lognormal 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. They in turn have given rise to a wide range of massive real-world social and information network data. The unifying theme is then to study real-world social and information networks, with an emphasis on finding and explaining patterns in large social networks, computer networks, Internet networks, communication networks, e-mail interactions, Web graphs, gene regulatory networks, and so on. The main aim 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?

Empirical observations and models that explain them. Research on networks has focused on two aspects. First, 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 [34], small diameter [3, 23], navigability [29, 17], and so on. Second, there have been efforts to find the network formation mechanisms that naturally capture such structural features. Again, there have been two relatively disjoint approaches to come up with desirable models. The approach mainly in the theoretical computer science has developed relatively simple but analytically tractable network models that naturally lead the network properties observed in the real-world. The prime examples of this line of work are following: the Preferential Attachment model and its variants [4, 1, 8, 9, 11, 14], the Copying Model [18], the Small-world model [34, 17], the Forest Fire model [24], the Random surfer model [5], and models of bipartite affiliation networks [19]. On the other hand, in statistics, machine learning and social network analysis, another approach has put efforts into 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. Such models include, the Exponential Random Graphs [32], the Stochastic Block Model [2] and the Latent Space Model [15].

"Mechanistic" and "Statistical" models. Generally, there has been some gap between these two lines of research. The emphasis with "mechanistic" models is on the analytical tractability in a sense not only that one can mathematically analyze properties of the networks that arise from the model but also that these network structures naturally emerge from the model. However, from the "statistical" point of view, such models are usually not interesting mainly due to their simplicity. On the contrary, most "statistical" models are analytically intractable and the network properties do not naturally emerge from the model in general. Conversely, such models are normally able to jointly model the node features as well as the network structure. They are usually accompanied by statistical procedures for model parameter estimation and are very useful for testing various hypotheses about the interaction of linking patterns and the properties of nodes.

However, models of network structure or formation are seldom *both* analytically tractable and statistically interesting. One example for models satisfying both features is the Kronecker graphs model [21, 35], 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 [28, 20, 6]. Besides, it 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 attributes. Network models in general investigate the relationships between nodes, but a rich set of attributes are associated with each node, especially in social networks. Traditionally, not only people's connections but also their characteristics, like age, gender, work place, habits, etc., have been collected as social network data via questionnaires with them.. Similarly, various profile information is provided by users in online social networks. In this sense, both node attributes and network structures need to be considered simultaneously in many situations.

The attempt to model both of them 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 hard to fit to real data or statistically very powerful but analytically intractable.

Here we propose a model that naturally captures the interactions between the network structure and node attributes in a clean and tractable manner. First, we consider a model where each node has a vector of categorical features, which are in turn characterized by their preferences to the similarities. For example, if people share certain features like hobby, they are more likely to be friends. Reversely, for some features like gender, people are more likely to have some relationships with the opposite characters. Each case is named as homophily(*i.e.* love of the same) and heterophily(*i.e.* love of the difference), respectively. Thus, the proposed model is designed to capture the homophily as well as the heterophily that naturally occurs in social networks. Furthermore, this preference can be naturally enhanced to the node level by multiplying all scores between a pair of nodes so that it represents the probability of the edge between them. Through the multiplicative way, the node attributes and the network structure can be nicely blended.

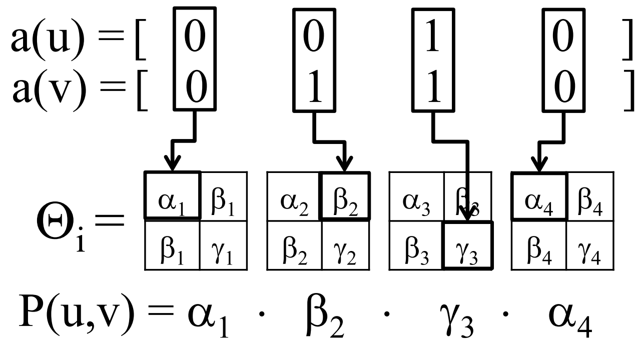

Figure 1: Schematic representation of the model. Given a pair of nodes u and v with the corresponding attribute vectors $a(u)$ and $a(v)$, the probability of edge $P(u, v)$ is the product over the entries of attribute-attribute similarity matrices Θ_i where values of $a_i(u)$ and $a_i(v)$ "select" the appropriate entry (row/column) of Θ_i.

We refer to this class of network models as the Multiplicative Attribute Graphs (MAG). We now proceed to formulate the model and show that it is both statistically interesting and mathematically tractable in the following.

1.1 Formulating Multiplicative Attributes Graph model

Since there are several aspects that the model should take account of, we undertake the Multiplicative Attributes Graph (MAG) model via a sequence of steps.

General considerations. The basic setting under which we operate is that each node u has a set $a(u)$ of l categorical attributes associated with it. For a simple example, we might assume that the attributes are binary. One can t hink of these attributes as if we ask weach node, *i.e.* each person in the social network, a series of l yes/no questions such as "Are you female?" or "Do you like ice creams?", and so on. A sequence of answers to these questions forms a binary vector of length l that associates with each node.

The second essential ingredient is to specify a mechanism that generates the probability of an edge between two nodes where the attribute vectors of the nodes are given. As mentioned before, we would like our model to be able to account for both the homophily of certain features and the heterophily of the others. To elaborate the idea mentioned before, the way that we here propose is to associate each feature i (*i.e.* i-th question) with an attribute-attribute similarity matrix Θ_i. For the above binary example, eacho Θ_i should be 2×2 matrix. The entries of matrix Θ_i represent the edge probability given the values of i-th attribute of both nodes. Thus, if the attribute reflects homophily, the corresponding matrix Θ_i would have large values on the diagonal (*i.e.* the probability of the edge is high when the nodes' answers match), while for attributes with heterophily the off-diagonal values of Θ_i would be high (*i.e.* the probability of the link is high when nodes gave different answers to the same question).

The Multiplicative Attributes Graph (MAG) model. Now we present the general version of the MAG model. First, let each node u have a vector of l categorical attributes associated with it. We assume that each attribute $i = \{1, \ldots, l\}$ has cardinality d_i. We also have a set of l matrices Θ_i of size $d_i \times d_i$. Each entry of Θ_i is a probability, *i.e.* a real value between 0 and 1. Then, as shortly mentioned before, the probability of an edge (u, v), $P(u, v)$, is defined as the multiplication of probabilities corresponding to individual attributes, *i.e.*

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

where $a_i(u)$ denotes the value of i-th attribute of node u. Notice 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 $a(u)$ associated with node u stores the answers of u to these questions. Θ_i then reflects the marginal probability of an edge over the answers of a pair of nodes for the i-th question. That is, 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 similarity matrices. Supposed that the questions are appropriately chosen so that answers should be independent, the product over the entries of matrices Θ_i results in the global probability of an edge between u and v.

The proposed model is statistically interesting as one can pose many attractive problems: given attribute vectors of all nodes and the network structure, how to estimate the values of matrices Θ_i or infer the attributes of unobserved nodes; or, given a network, how to estimate both the node attributes and the matrices Θ_i. A simple expectation maximization based method could be a approximate solution to estimate the model parameters, however, we leave the questions of the efficient model parameter estimation as the future work.

On the other hand, the proposed model is also mathematically tractable in a sense that we can formally analyze the properties of the model. However, since the model currently includes too many parameters to exhibit common properties, we next introduce a simplified version of the model.

Simplified version of the model. We here 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, for computational convenience we will consider only undirected networks that each Θ_i should be symmetric. Moreover, we assume binary attributes and thus matrices Θ_i have 2 rows/columns. 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. By all assumptions on Θ_i's, we can indeed set $\Theta = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix}$, *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 where $\alpha > \beta > \gamma$. As we show later, the assumption $\alpha > \beta > \gamma$ is very natural since most real-world networks have a common structure [20].

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

After all, 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, μ is the probability that attribute takes a value of 1 and $\Theta = [\alpha\ \beta; \beta\ \gamma]$ specifies the shared attribute-attribute similarity matrix.

We now study the properties of the random graph that results from the MAG model $M(n, l, \mu, \Theta)$ where every pair of nodes (u, v) is independently and undirectly connected with probability $P(u, v)$ defined in Equation 1. Since the probability exponentially decreases as a function of l, the most interesting case occurs when $l = \rho \log n$ for some constant ρ (Corollary 2.4).[1] Our analyses will focus on this case.

[1]Throughout the paper, every log notation indicates $\log_2$ unless we explicitly specify it as $\ln$.

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 attribute values 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 [36], the linking probability depends on the inner product between the vectors associated with node positions.

The MAG model generalizes the Kronecker graphs model [20] in a subtle and surprising 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 $2^l \times 2^l$, simply interpreted as the graph adjacency matrix. As shown in [20], one can think of a Kronecker graph model as a variant of the MAG model.

Proposition 1.1 *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 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. $P_G(u, v) = P_M(u, v)$.*

This 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 different combination of attribute values (*i.e.* each node has different node id) and all attribute value combinations are occupied (*i.e.* $0, \cdots, 2^l - 1$). Secondly, building on this correspondence between Kronecker and MAG graphs, we also note that the estimates of the Kronecker parameter matrix K nicely transfer to the MAG model. For example, $K = [\alpha = 0.98, \beta = 0.58, \gamma = 0.05]$ 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 matrix Θ would normally take. In this respect, the assumption that $\alpha > \beta > \gamma$ turns out to be completely 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 [25, 20]. In other words, the network is composed from denser and denser layers as one moves towards the core of the network. Basically, $\alpha > \beta > \gamma$ means that more edges are likely to appear between nodes which have value 1's on more attributes and these nodes form the core of the network. Since more edges appear between pairs of nodes with attribute value 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").

1.2 Summary of results

We study the following structural properties of networks that arise from the MAG model as a function of model parameters n, l, μ and Θ.

First of all, we show that the expected number of edges in $M(n, l, \mu, \Theta)$ scales as:

$$\frac{1}{2} n^{2+\rho \log\left(\mu^2 \alpha + 2\mu(1-\mu)\beta + (1-\mu)^2 \gamma\right)} \qquad \text{where } \rho = l / \log n \text{ (Theorem 2.3)}$$

Note that the logarithmic term is less than 0. Our model thus produces graphs that obey the Densification Power Law [23] which states that in real networks the number of edges $m(t)$ as a function of time t grows as $m(t) \propto n(t)^a$ for $a > 1$, *i.e.* average degree in the network increases over time.

Moreover, we study the conditions under which the MAG graph is connected (Theorem 4.1) and also show that the giant connect component of size $\Theta(n)$ emerges when:

$$\left[(\mu\alpha + (1-\mu)\beta)^{\mu} (\mu\beta + (1-\mu)\gamma)^{1-\mu}\right]^{\rho} \geq \frac{1}{2} \qquad \text{(Theorem 5.1)}$$

Besides, we show that a MAG graph has a constant diameter with high probability when $(\mu\beta + (1-\mu)\gamma)^{\rho} > \frac{1}{2}$ (Theorem 6.1).

Finally, we derive the analytical expression for the degree distribution which can have very different shapes depending on the parameter setting. However, under mild conditions, we can show that the degree distribution follows a lognormal (Section 3). We view this as particularly interesting because there has been a long standing debate about distinguishing power-law distributions from lognormal distributions in empirical data [30, 31] and what kinds of differences that this would make about the understanding of real-world networks. However, as we show later, our model can model networks with the degree distribution following a lognormal as well as those with power-law degree distributions.

Putting these results into context, we provide some normally expected values of model parameters that arise in real-world networks. As discussed above, we expect that $[\alpha = 0.98, \beta = 0.55, \gamma = 0.1]$ could be a typical example of parameter Θ. For instance, when we have a network with 1 million nodes (*i.e.* $n = 10^6$) and set $\mu = 0.5$ and $\rho = 0.75$, we expect to have roughly 30 million edges. This network will be just above the threshold for the emergence of the giant connect component, whereas it will be just below the threshold which would ensure that the network is connected. This example seems interesting since it hints that real-world networks are in the parameter region analogous to an extremely sparse Erdős-Rényi [12] G_{np} random graph model with $1/n < p < \log n/n$.

1.3 Extensions and further results

In addition to the simple version, we demonstrate the flexibility of the original model by showing how a simple and natural generalization from the simplified model can give rise to networks with power-law degree distributions with any power-law degree exponent $\delta > 1$. In order to obtain power-laws, all we need is to relax the assumption that all attributes are generated from the same Bernoulli distribution and share the similarity matrix. In other words, we assume that the i-th attribute is generated i.i.d. from a Bernoulli distribution with parameter μ_i, *i.e.* each attribute has a different probability of taking value of 1, as well as that the i-th similarity matrix is Θ_i. We then illustrate that for each power-law exponent δ there exists a setting of μ_i's and Θ_i's such that the power-law degree distribution arises.

There are clearly many directions in which the model could be generalized and improved. Our goal here is to explore some of the interesting phenomena that emerge already in a very simple version of the model, but one can easily generalize our framework to assume a different attribute generative models, or consider other ways to combine individual entries of Θ_i.

Another significant direction for the future lies in further understanding the properties of the model; for example, the degree of local clustering, spectral properties of the MAG model , or efficient (decentralized) searchability and navigability. Especially, we believe that the searchability problem is interesting since for an efficient local search algorithm one could use the information both about the node attributes and about the attribute-attribute similarity matrices. The setting in MAG model reminds us of that in [33] where authors consider that nodes of a tree reside in multiple places of a large hierarchy and the probability of connection depends on the tree distance between the positions of the corresponding nodes.

A very different direction of future work lies in parameter estimation of the MAG model. We here envision several versions of problem. The most general one is that we are given a network on n nodes and are asked to estimate attribute vectors of all nodes and the similarity matrices of all attributes. Other variants of the problem include the case where a network is missing some edges, *i.e.* some random set of edges is missing (unobserved, hidden) and the question is then to recover the missing edges based on the partial network. This is an interesting variant of the link prediction problem [26] where not only we have access to the network but also we know the attributes of nodes. Similarly, where given a network and partial node attribute values, one may be interested in the estimation of the missing attribute information. In such case, the actual objective would be to estimate the attribute similarity matrices based on a network structure and infer the missing attributes by using the similarity matrices as well as the network structure.

2 The Number of Edges

In this section, we deal with the expected number of edges in MAG model as its first property, which implies more importance than just counting edges. In order to compute the expectation, we equivalently study not only the expected degree of nodes but also the expected probability of connection between two random nodes, because these values are, thanks to the linearity of expectation, naturally followed by the total number of edges.

Furthermore, this mathematical analysis is followed by two substantial features. The first one is the validiation of the assumption, $l = \rho \log n$ for some constant ρ. One might easily make sure that the number of attributes cannot be too large since the edge probability would exponentially decrease as it grows in our model. Once the number of edges is expressed as a function of parameters including n and l, this expression can explain the assumption clearly and analytically. The other characteristic is the change of graph density when the number of nodes increases. As briefly mentioned before, social networks generally obey the Densification Power Law [23]. Under the $l = \rho \log n$ assumption, the number of edges can demonstrate that MAG model also nicely follows this law. These two features are studied at the end of this section.

Before the actual analysis, for convenience, we define some useful notations. First, let V be the set of nodes in the MAG graph M. We refer to the *weight* of a node as the number of 1's among its attributes. For a node $u \in V$, $|u|$ denotes its weight, *i.e.*

$$|u| = \sum_{i=1}^{l} \mathbf{1}\left\{a_i(u) = 1\right\}$$

where $\mathbf{1}\left\{\cdot\right\}$ is an indicator function. Additionally, W_j denotes a set which consists of 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 the definitions, $S_j = \cup_{i=j}^{l} W_i$ undoubtedly holds.

Since in the simplified MAG model each attribute is independently sampled from $Bernoulli(\mu)$, the weight of a node eventually follows the binomial distribution $Bin(l, \mu)$. The expected size of W_i is therefore easily computed as $\mathbb{E}\left[W_j\right] = n\binom{l}{j}\mu^j(1-\mu)^{l-j}$. With regard to $|S_j|$, it is the sum of contained W_i's because each W_i is disjoint with others.

Furthermore, we require another notation as follows. For a pair of nodes, u and v, $P[u, v]$ denotes the probability of connection between them where their attributes are provided. Note that it is conditioned on their attributes, which are also Bernoulli random variables. That is, $P[\cdot, \cdot]$ is a function of random variables, *i.e.* a random variable by itself. As a consequence, $\mathbb{E}\left[P\left[u, v\right]\right]$ represents the expected number of edges over the joint distribution of u and v's attributes.

On the other hand, since each connection is determined by Bernoull distribution, $P[u, v]$ is effectively the same as its expectation. In this sense, we can similarly define the expected connections between a node and a set of nodes or between two sets of nodes. Formally, for $S, S' \subset V$, $P[u,S]$ and $P[S,S']$ represent $\sum_{v \in S} P[u,v]$ and $\sum_{u \in S} P[u,S']$, respectively. Also both are undoubtedly random variables over the corresponding attributes.

As described above in brief, the expected number of edges is naturally led by the expected degree of a node, which can be obtained from the expected degree conditioned on the weight of the node. In other words, we first calculate $\mathbb{E}\left[P\left[u,V\right]|u \in W_i\right]$ and sum them up over the weight distribution. This conditional expectation can be quite simply derived from the following lemma:

Lemma 2.1 *For distinct $u, v \in V$,*

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

What this lemma represents is the conditional probability of connection where the weight of one node is given. Its rigorous proof is described in Appendix. By Lemma 2.1 and the linearity of expectation, we can sum this conditional probability over all nodes to result in the following lemma:

Lemma 2.2 *For $u \in V$, its expected degree is*

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

Since $|V \backslash u| = n-1$, the expected number of connections to other nodes is equal to the first term in Lemma 2.2. In contrast, the second term represents the double probability of the self-edge, effectively the expected degree of self-edges.

Since the number of edges is a half of the degree sum by definition, all we need to do is to average $\mathbb{E}\left[deg(u)|u \in W_i\right]$ over the *weight* distribution, *i.e.* binomial distribution $Bin(l,\mu)$ and halve it. Then, we can lead the expected number of edges as follows.

Theorem 2.3 *For the MAG graph $M(n,l,\mu,\Theta)$, the number of edges m is expected to be*

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

Proof:

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

■

This is also divided into two diffrent terms. The first term indicates the number of edges between different nodes, while the second term means the number of self-edges. Therefore, if we exclude self-edges, then the number of edges would be reduce to the first term. This term can be expressed in the vector form as $\binom{n}{2}\left(\mathbf{z}^T\Theta\mathbf{z}\right)^l$ where $\mathbf{z} = [\mu \;\; 1-\mu]^T$.

From Theorem 2.3, the density of graph turns out to be l-th power of an affine combination of α, β, and γ by definition. That is, as μ becomes larger, the expected probability goes to α^l and the graph becomes denser. Conversely, μ becomes smaller, then the probability decreases to γ^l and the graph becomes sparser. In this view, μ acts like a controller of the graph density.

However, more significantly, since the density of graph is l-th power of some constant less than 1, l cannot grows too large compared to n. Otherwise, the density drops so much that the graph cannot contain as many edges as nodes. Conversely, The following two corollaries describe it in detail.

Corollary 2.4 *If* $\frac{l}{\log n} > -\frac{1}{\log(\mu^2\alpha+2\mu(1-\mu)\beta+(1-\mu)^2\gamma)}$ *as* $n \to \infty$*, then* $m \in o(n)$ *with high probabiltiy.*

In other words, in order for $M(V, E)$ to have a proper number of edges, for example, more than n, l should be bounded by the order of $\log n$. We did not give any specific reason to set $l = \rho \log n$ before, but Corollary 2.4 provides it.

Corollary 2.4 is simply provable. Suppose that $l = \log n\,(\epsilon - 1/\log\zeta)$ for $\zeta = \mu^2\alpha + 2\mu(1-\mu)\beta + (1-\mu)^2\gamma$ and $\epsilon > 0$. By Theorem 2.3, the expected number of edges is $\Theta\left(n^2\zeta^l\right)$. However, since μ and γ is less than 1 so that $\zeta < 1$ and $\log\zeta < 0$, this expectation is after all equal to

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

On the contrary, one can think that l might be much less than $\log n$, *i.e.* $o(\log n)$. However, in this case, $M(V, E)$ becomes too dense, asymptotically close to $\Theta(n^2)$, as $n \to \infty$. Since most social networks are quite sparse, this case can be reasonably excluded. Therefore, both Corollary 2.4 and this exclusion is fully supportive of the assumption that $l = \rho \log n$.

Under this validated assumption, the expected number of edges without self-edges can be restated as follows:

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

We can easily confirm that this fact agrees with the Densification Power Law[23] that $m(t) \propto n(t)^a$ for $a > 1$. For example, Kronecker graph, an instance of $\rho = 1, \mu = 0.5$ (Proposition 1.1), would have the desenification exponent $a = \log(|\Theta|)$ where $|\Theta|$ denotes the sum of all entries in Θ.

Although we do not define any process here, we can interpret it in the folllowing way. When a new node comes in, its behavior is governed by the node distribution seemingly independent of the graph structure. However, in the long term, since the number of attributes grows slowly as the number of nodes increaes, they are not obviously independent. This phenomenon is somewhat aligned with the real world. When a new person enters the network, he or she seems to act independently, but people eventually constitue a structured network in the large scale and their behaviors can be categorized into more classes as the network evolves. Thus, this simplified model implies that the network might evolve in the $\log n$ order.

3 Degree Distribution

In this section, we would like to capture the degree distribution for the simplified MAG modelunder some reasonable assumptions.[2] Otherwise, its shape might be very different relying on the parameter setting.

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

For instance, since the graph almost becomes a (sparse) Erdös-Rényi random graphif $\alpha \approx \beta \approx \gamma < 1$, the degree distribution will approximately follow the binomial distribution. For another extreme example, in case of $\alpha, \mu \approx 1$, the graph will be close to the clique, which obviously represents a different degree distribution from a sparse Erdös-Rényi random graph.

For this reason, we have to narrow down the condition on μ and Θ as follows. First, μ close to 0 or 1 nearly leads a Erdös-Rényi random graph with edge probability $p = \alpha(\mu \approx 1)$ or $\gamma(\mu \approx 0)$. Since the degree distribution Erdös-Rényi random graph is trivially binomial, we will exclude this extereme case of μ. On the other hand, with regard to Θ, the differences not only between α and β but also between β and γ seem pretty large from the experimental results in [22]. We thus guess that a reasonable configuration space for Θ would be placed where $\frac{\mu\alpha+(1-\mu)\beta}{\mu\beta+(1-\mu)\gamma}$ is $1.6 \sim 3$. For the previous Kronecker graph example, its ratio is actually about 2.44. Moreover, we figure out that the minimum ratio of examples in [22] is around 1.7. Our approach for these additional conditions could be therefore supported by those real examples. Additionally, because we want to study normal real-world networks which usually have giant components, the analysis also assumes that the giant component exists, *i.e.*

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

The detailed explanation of this giant component condition is independently described in Section 5.

Under these assumptions, we are able to prove that for sufficiently large n the degree distribution approximately illustrates the quadratic relationship in the log-log scale like log-normal distribution. This result is totally acceptable since some social networks follow the log-normal distribution. For instance, the *Live-Journal* social network appears more parabolic than linear in the degree distribution[27].

In brief, since the average degree is a function of the node weight by Lemma 2.2, we can imagine that the degree distribution might be mainly affected by the distribution of node *weight*. This node weight actually follows binomial, however, when l is sufficiently large, it can be approximated to the normal distribution. This normality of node weights will eventually lead the log-normality of degrees. In the rest of this section, we show how the log-normality is derived.

Since the attributes of nodes are independent each other, the following general degree distribution formula[36] would nicely work:

$$P\left(deg(u) = k\right) = \int_{u \in V} \binom{n-1}{k} \left(\mathbb{E}\left[P\left[u, v\right]\right]\right)^{k} \left(1 - \mathbb{E}\left[P\left[u, v\right]\right]\right)^{n-1-k} du \tag{2}$$

If we assign our expectations (Lemma 2.1) into the above equation, then

$$P\left(deg(u) = k\right) = \sum_{j=0}^{l} \binom{l}{j} \mu^{j} (1-\mu)^{l-j} f_j(k) \tag{3}$$

where $f_j(k) = \binom{n-1}{k} \left(x^j y^{l-j}\right)^k \left(1 - x^j y^{l-j}\right)^{n-1-k}$ for $x = (\mu\alpha + (1-\mu)\beta)$ and $y = (\mu\beta + (1-\mu)\gamma)$.

Unfortunately, since it seems difficult to find the exact closed form of Equation 3, we hope to approximate this equation to some familiar closed form like polynomial. In order to employ several approximations, we require a couple of assumptions on the degree k. First, because the degree of maximum weight node is expected to be $O(n\left(\mu\alpha + (1-\mu)\beta\right)^{l})$, k would be $o(n)$ with high probability. Second, as we can expect that the median weight of nodes would be roughly μl by Central Limit Theorem, almost a half of nodes have degrees more than $(\mu\alpha + (1-\mu)\beta)^{\mu l} \left(\mu\beta + (1-\mu)\gamma\right)^{(1-\mu)l}$, *i.e.*the expected degree of median weight node (Lemma 2.2). Therefore, the assumption that $k > (\mu\alpha + (1-\mu)\beta)^{\mu l} (\mu\beta + (1-\mu)\gamma)^{(1-\mu)l}$ is totally acceptable when we focus on the tail of distribution.

Now we can consider Equation 3 as it is summed up by terms corresponding to j. Fortunately, most of those terms turn out to be insignificant under our assumptions. Then, the probability $p_k = P(deg(u) = k)$ is approximately proportional to one or few dominant terms. We thus seek for j such that maximizes $g_j(k) = \binom{l}{j}\mu^j(1-\mu)^{l-j} f_j(k)$ and show that $\log g_j(k)$ is a quadratic function of $\log k$

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

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

However, due to $k \in o(n)$,

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

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

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

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

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

for a constant C'.

Considering $g_j(k)$ as a function of j, not k, the following lemma describes j such that mazimizes the $g_j(k)$.

Lemma 3.1 $\arg\max_j g_j(k) \approx \tau$ *where MAG graph M has a giant component and $\tau \ge \mu l$.*

It is proved is in Appendix. As described before and explained in Section 5, the existence of the giant component is a sufficient and necessary condition that $\left[(\mu\alpha + (1-\mu)\beta)^{\mu} (\mu\beta + (1-\mu)\gamma)^{1-\mu}\right]^{\rho} \ge \frac{1}{2}$. By this relationship, since $\tau \ge \mu l$ and $l = \rho \log n$, k is much greater than the first quadratic term. Then, when $\left(\frac{x}{y}\right)$ is practical (close to $1.6 \sim 3$), $\ln g_{\tau+\Delta}$ would be at most $-ck|\Delta| \ln g_\tau$ for a constant c. After all, g_τ effectively dominates the probability p_k, *i.e.* $\ln p_k$ is roughly proportional to $\ln g_\tau$. By assiging $\tau = \frac{\ln k - \ln n y^l}{\ln\left(\frac{x}{y}\right)}$, we are able to obtain

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

which is a quadratic function of $\ln k$. Therefore, we can conclude that the degree distribution roughly follows the log-normal.

In the log-log plot, the slope varies from $\frac{1}{2}$ to $\left(\frac{1-\mu}{\mu}+\frac{1}{2}\right)$ when τ changes from μl to l. Thus, at the tail of distribution, the power-law might be observed when $\left(\frac{1-\mu}{\mu}\right)$ is comparatively small.

4 Connectivity

In this section, we here want to seek for the threshold for the connectivity of MAG model, one of the useful graph property, as done in [28]. Here is the overview of the connectivity. Since the node of less weight is more likely to be isolated, we find out the node of minimum weight in the graph to ensure that it is connected to the core. This minimum weight is a random variable, but its ratio to l converges to some constant which varies according to μ. We in turn search for the lower bound for the expected degree of the minimum weight node so that the graph should be connected. The combination of two assertions can consequently result in the following theorem for the criteria of the connectivity as a function of n, l, μ, and Θ[3].

Theorem 4.1 (Connectivity) *As $n \to \infty$, MAG graph M is connected with high proabilibty if*

$$\begin{cases} (\mu\beta + (1-\mu)\gamma)^\rho > \frac{1}{2} & \textit{when } (1-\mu)^\rho \geq \frac{1}{2} \\ \left[(\mu\alpha + (1-\mu)\beta)^\nu \left(\mu\beta + (1-\mu)\gamma\right)^{1-\nu}\right]^\rho > \frac{1}{2} & \textit{otherwise} \end{cases}$$

In contrast, M is disconnected with high probability if

$$\begin{cases} (\mu\beta + (1-\mu)\gamma)^\rho < \frac{1}{2} & \textit{when } (1-\mu)^\rho \geq \frac{1}{2} \\ \left[(\mu\alpha + (1-\mu)\beta)^\nu \left(\mu\beta + (1-\mu)\gamma\right)^{1-\nu}\right]^\rho < \frac{1}{2} & \textit{otherwise} \end{cases}$$

where $0 < \nu < \mu$ is a solution of the following equation:

$$\left[\left(\frac{\mu}{\nu}\right)^\nu \left(\frac{1-\mu}{1-\nu}\right)^{1-\nu}\right]^\rho = \frac{1}{2} \tag{4}$$

Note that the criteria are separated into two cases under the condition on μ. If we take a deep look at this condition, we are able to easily figure out what it means. This condition tells us whether the expected number of weight 0 nodes, *i.e.* $\mathbb{E}\left[|W_0|\right]$, is greater than 1 or not, 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.* $O(1)$. In the other case where the expected number of weight 0 nodes is quite small, we need to probe the minimum weight approximately. Equation 4 describes the ratio of the minimum weight to l as n goes to infinity.

To state the proof briefly, the disconnectivity can be proved by showing that the expected degree of the minimum weight node is too small to be connected with other nodes. Conversely, if the expected minimum degree is large enough, say $\Omega(\log n)$, then any subset of nodes will be connected with the other part of the graph. However, to reach this claim, we essentially need the following theorem.

Theorem 4.2 (Monotonicity) *For $u, v \in V$,*

$$\mathbb{E}\left[P\left[u,v\right] \mid |u| = i\right] \leq \mathbb{E}\left[P\left[u,v\right] \mid |u| = j\right] \quad \text{if} \quad i \leq j$$

[3]We assume that $\rho = n/\log n$

Theorem 4.2 demonstrates that the node of more weight is more likely to be connected with the others. That is, as the weight of a node is larger, the degree of it tends to be higher. In another view, as mentioned before, the high degree nodes play "core" roles in the graph, whereas the low degree nodes become "peripheries". This feature of MAG model has great effects on the connectivity as well as on the existence of giant component (Section 5).

By the way, we can trivially derive the following corollary that will play a main role in showing the lower bound of degree for the connectivity.

Corollary 4.3 $\mathbb{E}\left[P\left[u,v\right]|v \in S_j\right] \geq \mathbb{E}\left[P\left[u,v\right]|v \in S_i\right]$

This might be obvious since the (weighted) average of larger values is greater than that of smaller value. The mathematical proofs of both theorem and corollary are described in Appendix.

Now we seek for the minimum weight of the graph. Let the minimum weight in V be $V_{\min}$. Undoubtedly, it has some distribution depending on the parameters, μ and Θ. The following two lemmas illustrate the level of the minimum depending on the situations.

Lemma 4.4 *If* $(1-\mu)^\rho \geq \frac{1}{2}$, *then* $V_{\min} \in O(1)$ *with high probability as* $n \to \infty$.

Lemma 4.5 *If* $(1-\mu)^\rho < \frac{1}{2}$ *and* $\nu < \mu$ *is a solution of the following equation:*

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

, then $\frac{V_{\min}}{l} \to \nu$ *with high probability as* $n \to \infty$.

The reason for the separation is exactly the same as in Theorem 4.1. Their proofs are described in Appendix.

Now we move onto the degree part of proof assuming that the above two lemmas hold.

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

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

Since the probability that there exists no edge between S' and $S\backslash S'$ is at most $\exp\left(-\frac{1}{2}\mathbb{E}\left[P\left[S', S\backslash S'\right]\right]\right)$ by Chernoff bound, it is bounded as follows:

$$\begin{aligned}
P(S \text{ is disconnected}) &\leq \sum_{S' \subset S} P(\text{no edge between } S', S\backslash S') \\
&\leq \sum_{S' \subset S} \exp\left(-\frac{c\log n}{2n}|S'| \cdot |S\backslash S'|\right) \\
&\leq 2\sum_{i \leq |S|/2} \binom{|S|}{i} \exp\left(-\frac{c|S|\log n}{2n} i\right) \\
&\leq 2\sum_{i \leq |S|/2} \exp\left(\left(\log|S| - \frac{c|S|\log n}{2n}\right) i\right) \in o(1)
\end{aligned}$$

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

Let $\frac{V_{\min}}{l} \to t$ for a constant $0 \le t < \mu$ as $n \to \infty$. Then, this setting can cover both cases of Lemma 4.4 and Lemma 4.5.

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

$$\mathbb{E}\left[P\left[u, V \backslash u\right] | u \in W_{V_{\min}}\right] \approx \mathbb{E}\left[P\left[u, V \backslash u\right] | u \in W_t\right] \approx \left[2\left[(\mu\alpha + (1-\mu)\beta)^t (\mu\beta + (1-\mu)\gamma)^{1-t}\right]^\rho\right]^{\log n}$$

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

On the other hand, when $(\mu\alpha + (1-\mu)\beta)^{\frac{V_{\min}}{\log n}} (\mu\beta + (1-\mu)\gamma)^{\frac{l-V_{\min}}{\log n}} < 1$, 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.

5 Giant Connected Component

With regard to the global properties of social networks, the existence of the giant connected component is more general than the connectivity. We here generalize the idea in [28] to induce the sufficient and necessary condition for MAG model in the following theorem.

Theorem 5.1 (Connected Component) *The sufficient and necessary condition for the existence of the giant component with size $\Theta(n)$ in M is*

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

where $n \to \infty$.

The similar explanation to the connectivity would work here. Theorem 5.1 describes that the existence of $\Theta(n)$ component relies on the degree of the *median* weight, instead of the minimum weight. The median degree nodes usually play "periphery" roles in the giant component rather than low degree nodes.

We can consider this in the following way. Assume that we randomly choose a pair of nodes from the connected graph and remove their connection if exists. As the graph becomes sparse, the low degree nodes tend to be isolated from the main component. After almost half of nodes are separated from the core, the nodes of median degree now tends to have few connections. Even though the graph can be much sparser, it still includes the giant component of a half size.

The proof consists of two major parts: one is the proof/disproof of existence of $\Theta(n)$ component, and the other is the proof of uniqueness. When proving/disproving the existence, depending on certain conditions, we will suggest the corresponding examples/counter-examples. However, the existence of $\Theta(n)$ component is not the sufficient and necessary condition that it is a giant component, since there might be another $\Theta(n)$ component. Therefore, to prove it more strictly, the uniqueness of $\Theta(n)$ component has to follow the existence of it. For this uniqueness, we will just show that if there are two connected components of size $\Theta(n)$ then they are connected each other with high probability.

We now find the conditions where the component of size $\Theta(n)$ exists. Specifically, we check the three kinds of subgraphs: $S_{\mu l}$, $S_{\mu l + l^{2/3}}$, and $S_{\mu l + l^{1/6}}$ [28]. The following lemmas tell us the size of each subgraph.

Lemma 5.2 $|S_{\mu l}| \ge \frac{n}{2} - o(n)$ *as* $n \to \infty$.

Lemma 5.3 $|S_{\mu l+l^{2/3}}| = o(n)$ *as* $n \to \infty$.

Lemma 5.4 $|S_{\mu l+l^{1/6}}| \in \Theta(n)$ *as* $n \to \infty$.

Rigorous proofs of these are described in Appendix.

We in turn consider each condition that $\left[(\mu\alpha + (1-\mu)\beta)^{\mu}(\mu\beta + (1-\mu)\gamma)^{1-\mu}\right]^{\rho}$ is larger than, equal to, or less than $\frac{1}{2}$.

First, if $\left[(\mu\alpha + (1-\mu)\beta)^{\mu}(\mu\beta + (1-\mu)\gamma)^{1-\mu}\right]^{\rho} > \frac{1}{2}$, then since $|S_{\mu l}| \in \Theta(n)$ by Lemma 5.2, $S_{\mu l}$ is connected with high probability by Section 4. 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 because

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

is greater than $c\log n$ and $|S_{\mu l+l^{1/6}}| \in \Theta(n)$. Thus, $S_{\mu l+l^{1/6}}$ is connected with high probability.

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

$$\mathbb{E}\left[P\left[u, V\backslash u\right]\right] \approx \left[\left[(\mu\alpha + (1-\mu)\beta)^{\mu}(\mu\beta + (1-\mu)\gamma)^{1-\mu}\right]^{\rho}\right]^{\log n}\left(\frac{(\mu\alpha + (1-\mu)\beta)}{(\mu\beta + (1-\mu)\gamma)}\right)^{(\rho\log n)^{2/3}}$$

is $o(1)$ as $n \to \infty$. Since $S_{\mu l+l^{2/3}}$ is $o(n)$ by Lemma 5.3, most of $n - o(n)$ nodes are isolated, therefore the size of the largest component cannot be $\Theta(n)$.

We now check the uniqueness of the $\Theta(n)$ connected component.

We already pointed out that $S_{\mu l}$ or $S_{\mu l+l^{1/6}}$ is included in the $\Theta(n)$ component. Let that component be H. Then, since $\mathbb{E}\left[P\left[u,v\right]|v \in H\right] \geq \mathbb{E}\left[P\left[u,v\right]|v \in V\backslash H\right]$ by Theorem 4.2, $\mathbb{E}\left[P\left[u, V\backslash H\right]\right] \leq \frac{n-|H|}{|H|}\mathbb{E}\left[P\left[u,H\right]\right]$ holds for every $u \in V$.

Assume that another connected component H' also contains $\Theta(n)$ nodes. As described before, we hope to show that it is connected to H with high probability to present the contradiction.

$$\begin{aligned}
\mathbb{E}\left[P\left[H,H'\right]\right] &= |H'|\mathbb{E}\left[P\left[u,H\right]|u \in H'\right] \\
&\geq \frac{|H'|\cdot|H|}{n-|H|}\mathbb{E}\left[P\left[u,V\backslash H\right]|u \in H'\right] \\
&\geq \frac{|H'|\cdot|H|}{n-|H|}\mathbb{E}\left[P\left[u,H'\right]|u \in H'\right]
\end{aligned}$$

Since both $|H|$ and $|H'|$ are $\Theta(n)$, if $\mathbb{E}\left[P\left[u,H'\right]|u \in H'\right]$ is $\Omega(1)$, then $\mathbb{E}\left[P\left[H,H'\right]\right] \in \Omega(n)$, which indicates that H and H' is connected with high probability.

On the other hand, if $\mathbb{E}\left[P\left[H,H'\right]\right]$ is $o(n)$, i.e., $\mathbb{E}\left[P\left[u,H'\right]|u \in H'\right] \in o(1)$, 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.

6 Diameter

Another well-known property of social networks is that the diameter of the network remains constant even though the number of nodes is huge. By introducing main ideas in [28], we can also show this feature in MAG model as follows:

Theorem 6.1 (Constant Diameter) *If $(\mu\beta + (1-\mu)\gamma)^\rho > \frac{1}{2}$, then MAG model has a constant diameter with high probability as $n \to \infty$.*

This theorem does not specify the exact number of diameter, but, at least, it guarantees the constant diameter by offering the constant upper bound under the given conditions including that n is sufficiently large.

We begin the proof with the introduction of an important lemma to hint the upper bound for the diameter.

Lemma 6.2 *[28, 7, 16] For a 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 goes to infinity.*

Lemma 6.2 describes only when the graph is a Erdös-Rényi random graph. However, if we can assure that the probability between any pair of nodes is greater than p, then it is obvious that the diamter of the graph would be at most that of $G(n,p)$ [28].

To show a constant diameter, we will propose a subgraph such that each edge probability is greater than that of Erdös-Rényi random graph described in Lemma 6.2. By definition, this proposed subgraph has a constant diameter. If this subgraph is additionally directly connected with the rest part of the graph, then we are able to conclude the constant diameter of the entire graph.

We now suggest $S_{\lambda l}$ for $\lambda = \frac{\mu\beta}{\mu\beta+(1-\mu)\gamma}$ and prove that this subgraph satisfies the above properties where $(\mu\beta + (1-\mu)\gamma)^\rho > \frac{1}{2}$.

First, show that $S_{\lambda l}$ has a constant diamter with high probability. To do so, since $\min_{u,v\in S_{\lambda l}} P[u,v] \geq \beta^{\lambda l}\gamma^{(1-\lambda)l}$, it is enough to prove that $G(|S_{\lambda l}|, \beta^{\lambda l}\gamma^{(1-\lambda)l})$ has a constant diameter.

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

Since $|W_{\lambda l}|$ goes to $\mathbb{E}\left[W_{\lambda l}\right]$ 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 Lemma 6.2, a Erdös-Rényi random graph $G(|S_{\lambda l}|, \frac{c(1+\epsilon)^{\log n}}{|S_{\lambda l}|\sqrt{l}})$ has diameter at most $\left(1+\frac{1}{\epsilon}\right)$ as $n \to \infty$. Thus, the diamter of $S_{\lambda l}$ is also bounded by a constant.

Second, we need to show that every node in $V \backslash S_{\lambda l}$ is directly connected to $S_{\lambda l}$ with high probability. For any $u \in V$,

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

By Centeral Limit Theorem, $\sum_{j=\lambda l}^{l} \lambda^j (1-\lambda)^{l-j}$ is approximately $\frac{1}{2}$. Therefore, $\mathbb{E}\left[P\left[u, S_{\lambda l}\right]\right]$ 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.

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 observed. However, as for the degree distribution, even though the log-normal is one of the distributions that social networks could follow, a lot of social networks are known to follow the power-law degree distribution. Unforunately, in the simplified MAG model , we are not able to capture this property.

In this section, we show that MAG model might produce the desired property by releasing some constraints. We do not attempt to analyze it in a rigorous manner, but give the intuition about the model by suggesting an example of configuration.

We now propse the model that follows the power-law degree distribution by increasing the number of parameters. We still hold the condition that every attribute is binary and independently sampled from Bernoulli distribution. In contrast to the simple version, we do not equalize the attribute distributions as well as the similarity matrices. The formal definition of relaxed model is as follows:

$$P\left(a_j(u) = 1\right) = \mu_j$$

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

After all, the number of parameters increases up to $(4 \times l)$, which consist of μ_j's and Θ_j's for $j = 1, 2, \cdots, l$. With these more parameters, we are actually able to obtain the approximate power law degree, $\log p_k \propto k^{-\delta}$, for any power $\delta > 1$.

Let the ordered probability masses of attributes events be $p_{(j)}$ for $j = 1, 2, \cdots, 2^l$. For example, if the probability of each attributes event $(00, 01, 10, 11)$ 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 Equation 2, the probability of degree k, p_k is as follows:

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

where E_j denotes the expected edge probability of attributes corresponding to $p_{(j)}$. If E_j's are spread out as in Section 3 so that few terms dominate the probability, we might approximate

$$\ln p_k \approx \ln \binom{n-1}{k} p_{(\tau)} (E_\tau)^k (1 - E_\tau)^{n-1-k}$$

for τ such that $E_\tau \approx \frac{k}{n}$.

Applying the same algebra in Seciont 3, eventually the following statement would hold:

$$\ln p_k \approx C + \ln p_{(\tau)} - \frac{1}{2}\ln k$$

Therefore, if we could configure as

$$p_{(\tau)} \propto k^{-(\delta-\frac{1}{2})} \propto E_\tau^{-(\delta-\frac{1}{2})} \tag{5}$$

then it will follow the power-law degree distribution with power δ. We give an example that satisfies this condition. To show that, we need the following lemma.

Lemma 7.1 *The expected edge probability of this extended model is*

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

This lemma can be easily proved by the mathematical induction.

On the other hand, the probability mass for the attributes is

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

From these these two probability formula, if every $p_{(j)}$ is distinct and $\frac{\mu_i}{1-\mu_i} \propto \left(\frac{\mu_i\alpha_i+(1-\mu_i)\beta_i}{\mu_i\beta_i+(1-\mu_i)\gamma_i}\right)^{-(\delta-\frac{1}{2})}$, then we are able to satisfy the relationship in (5). Since Θ_i's are free to be configured, the above setting is totally feasible.

One thing that we should be careful of is to make every $p_{(j)}$ different. It is also a tractable condition. For example, if we set $\frac{\mu_i}{1-\mu_i} \propto \epsilon^{2^i}$, then they would work well.

References

[1] W. Aiello, F. Chung, and L. Lu. A random graph model for massive graphs. In *STOC '00: Proceedings of the 32nd annual ACM symposium on Theory of computing*, pages 171–180, 2000.

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

[3] R. Albert, H. Jeong, and A.-L. Barabási. Diameter of the world-wide web. *Nature*, 401:130–131, September 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: Proceedings of the 3rd Workshop on Analytic Algorithmics and Combinatorics*, 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] 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: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, 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: Proceedings of the 9th international conference on World Wide Web*, 2000.

[11] C. Cooper and A. Frieze. A general model of web graphs. *Random Structures and Algorithms*, 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 Acadamy 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: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, pages 251–262, 1999.

[14] A. D. Flaxman, A. M. Frieze, and J. Vera. A geometric preferential attachment model of networks. In *WAW '04: Proceedings of the 2nd Workshop On Algorithms And Models For The Web-Graph*, 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: Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, page 57, 2000.

[19] S. Lattanzi and D. Sivakumar. Affiliation networks. In *STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 427–434, New York, NY, USA, 2009. ACM.

[20] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani. Kronecker Graphs: An Approach to Modeling Networks. *Journal of Machine Learning Research*, 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: Proceedings of the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases*, pages 133–145, 2005.

[22] J. Leskovec and C. Faloutsos. Scalable modeling of real graphs using kronecker multiplication. In *ICML '07: Proceedings of the 24th International Conference on Machine Learning*, 2007.

[23] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shirinking diameters. *ACM Transactions on Knowledge Discovery from Data*, 2007.

[24] J. Leskovec, J. M. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *KDD '05: Proceeding of the 11th ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 177–187, 2005.

[25] 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: Proceedings of the 17th International Conference on World Wide Web*, 2008.

[26] D. Liben-Nowell and J. M. Kleinberg. The link prediction problem for social networks. In *12th CIKM*, pages 556–559, 2003.

[27] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. Geographic routing in social networks. *Proceedings of the National Academy of Sciences*, 102(33):11623–11628, Aug 2005.

[28] M. Mahdian and Y. Xu. Stochastic kronecker graphs. Technical report, 5th Workshop on Algorithms and Models for the Web-Graph (WAW), 2007.

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

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

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

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

[33] D. J. Watts, P. S. Dodds, and M. E. J. Newman. Identity and search in social networks. *Science*, 296:1302–1305, 2002.

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

[35] P. M. Weichsel. The kronecker product of graphs. In *American Mathematical Society*, volume 13, pages 37–52, 1962.

[36] 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

Proof of Lemma 2.1: Let N^0_{uv} be the number of 0-attributes shared by u and v. In a similar way, we can define N^1_{uv} as the number of 1-attributes shared. Clearly, $N^0_{uv}, N^1_{uv} \geq 0$ and $N^0_{uv} + N^1_{uv} \leq l$ hold. Then, by the definition of MAG model, the probability of connection between u and v is $\alpha^{N^1_{uv}} \beta^{l-N^0_{uv}-N^1_{uv}} \gamma^{N^0_{uv}}$.

However, N_{uv}^j's are also random variables functioned by attributes. Now, we find out the conditional joint distribution of N_{uv}^j's where the weight of u is equal to i. Each attribute is independently 1 with probability μ and 0 with probability $1-\mu$. That is, N_{uv}^0 and N_{uv}^1 are also independent each other. Eventually, N_{uv}^0 represents the number of 0-attributes in v among $(l-i)$ 0-attributes in u. It is exactly equal to the number of heads in $(l-i)$ coin tosses with probabiltiy $(1-\mu)$, which follows $Bin(l-i, 1-\mu)$. Similarly, N_{uv}^1 follows $Bin(i, \mu)$. Therefore, its 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 u's weight:

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

■

Proof of Lemma 3.1: By Theorem 5.1, the existence of a giant component indicates that

$$k \geq \left[\left(\mu\alpha + (1-\mu)\beta\right)^{\mu} \left(\mu\beta + (1-\mu)\gamma\right)^{1-\mu}\right]^{\rho} >> l$$

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

$$(\ln g_j(k))' \approx -\frac{j-\tau}{l\mu(1-\mu)} - \left(k \ln\left(\frac{x}{y}\right) - \frac{\tau - \mu l}{l\mu(1-\mu)}\right) - k\left(\frac{x}{y}\right)^{j-\tau} \ln\left(\frac{x}{y}\right) = 0$$

Since k is large enough and $(j-\tau) \in O(l)$,

$$\ln\left(\frac{x}{y}\right) \approx \left(\frac{x}{y}\right)^{j-\tau}$$

Therefore, when j is close enough to τ, $g_j(k)$ is maximized. ■

Proof of Theorem 4.2: For any $v_i \in W_i$, we can easily find a node v_i' by flipping $j-i$ zero bits randomly in v_i so that $P[u, v_i'] \geq P[u, v_i]$. Then, it is obvious that $\mathbb{E}\left[P\left[u, v_i'\right]|v_i\right] \geq \mathbb{E}\left[P\left[u, v_i\right]\right]$. Therefore,

$$\begin{aligned}
\mathbb{E}\left[P\left[u,v\right]|v \in W_j\right] &= \mathbb{E}\left[\mathbb{E}\left[P\left[u, v_i'\right]|v_i\right]\right] \\
&\geq \mathbb{E}\left[\mathbb{E}\left[P\left[u, v_i\right]|v_i\right]\right] \\
&= \mathbb{E}\left[P\left[u,v\right]|v \in W_i\right]
\end{aligned}$$

■

Proof of Corollary 4.3:

$$
\begin{aligned}
\mathbb{E}\left[P\left[u,v\right]|v\in S_j\right] &= \sum_{k=j}^{l} P(v\in W_k|v\in S_j)\mathbb{E}\left[P\left[u,v\right]|v\in W_k\right] \\
&\geq \sum_{k=j}^{l} P(v\in W_k|v\in S_i)\mathbb{E}\left[P\left[u,v\right]|v\in W_k\right] \\
&\quad + \sum_{k=i}^{j} P(v\in W_k|v\in S_i)\mathbb{E}\left[P\left[u,v\right]|v\in W_k\right] \\
&= \mathbb{E}\left[P\left[u,v\right]|v\in S_i\right]
\end{aligned}
$$

■

Proof of Lemma 4.4: The probability that $|W_c| = 0$ is at most $\exp(-\frac{1}{2}\mathbb{E}\left[|W_c|\right])$ by Chernoff bound. Since $\mathbb{E}\left[|W_c|\right] \geq \frac{n(l-c)^c}{c!}\mu^c(1-\mu)^{l-c}$, its probability goes to zero as l and n increase for fixed μ. Therefore, for sufficiently large l and n, $V_{\min}$ should be at most c with high probability. ■

Proof of Lemma 4.5: For any $\mu - \nu > \epsilon > 0$,

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

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

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

for a 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.

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. Thus, the expected number of nodes with at most weight $(\nu-\epsilon)l$ is less than $(\nu-\epsilon)l\ \mathbb{E}\left[|W_{(\nu-\epsilon)l|}\right]$ so that it also drops to zero as l increases. 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 νl with high probability. ■

Proof of Lemma 5.2: By Central Limit Theorem, $(|u| - \mu l) \sim \sqrt{l\mu(1-\mu)}N(0,1)$ as $n \to \infty$. Therefore, $P(|u| \geq \mu l)$ is at least $\frac{1}{2} - o(1)$. Then, by the Law of Large Number,

$$|S_{\mu l}| \to \frac{n}{2}$$

as $n \to \infty$. ■

Proof of Lemma 5.3: By Chernoff bound, $P(|u| \geq \mu l + l^{2/3})$ is $o(1)$ as $n \to \infty$, thus $|S_{\mu l + l^{2/3}}|$ is $o(n)$ with high probability. ■

Proof of Lemma 5.4: By Central Limit Theorem mentioned in Lemma 5.2,

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

is $o(1)$ as $n \to \infty$ 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 $\Theta(n)$. ■