Modeling Graphs with Attributes*

Neal Parikh
Department of Computer Science
Stanford University
npparikh@cs.stanford.edu

ABSTRACT

It is well known that real-world networks from many domains exhibit a large number of structural similarities, such as the form of the distributions of node degrees, the spectrum of the adjacency matrix, and the degree of interconnectedness. Some of these network statistics, such as the clustering coefficient, capture local behavior, while others, like the diameter, capture more global structure. After it was shown that many well-studied classes of stochastic network models, like Erdös-Renyi random graphs, do not accurately capture most of these properties, many researchers have proposed generative models that attempt to capture various aspects of network structure. The *stochastic Kronecker graph* model has been promising because it appears to capture many different properties reasonably well, rather than capturing one at the expense of the others, and because efficient algorithms exist for fitting the model to massive datasets. However, due to its rigid construction, it is difficult to modify or to extend to take advantage of richer datasets, such as node or edge-level attributes that may be available. This paper introduces and investigates *binomial attribute graphs*, a new model inspired by the Kronecker graph construction. Surprisingly, despite discarding a large amount of structure imposed by the Kronecker graph construction, the new model performs at least as well, and is much more easily modified, as it follows standard probabilistic semantics. Binomial attribute graphs are too simplistic to serve as general-purpose network models, but they provide new insight into why the Kronecker graph construction may work and suggest directions for improvement.

1. INTRODUCTION

The recent availability of large, real-world datasets encoding complex networks has led to a surge of research studying networks and their properties. A particular focus has been on the computational analysis of social networks, partly due to the availability of very large datasets in this domain. One surprising theme in this research has been the identification of of recurring structural features in a myriad of very different networks, including static features like power-law degree distributions, small diameter (the "small world phenomenon"), and skewness in the eigenvalues of the graph adjacency matrix; as well as temporal features like shrinking diameters and densification power laws. The fact that the same characteristics appear again and again in different domains has prompted an interest in network models, which propose generative processes for network structure and hope to uncover the fundamental processes that give rise to these global characteristics.

The earliest work in this domain dates back to the seminal paper of Erdös and Renyi on random graphs [3], in which each pair of nodes has an independent, identical probability p of being joined by an edge. Though this model has led to a rich mathematical theory, the graphs produced by this model do not match the empirical features of real-world datasets. For example, random graphs necessarily have a binomial degree distribution, while real-world networks exhibit power-law behavior. Since then, researchers in various fields have proposed a variety of formal models that attempt to capture characteristics of real networks. One idea that has dominated the literature is that of *preferential attachment* [1], in which a new node u links to a node v with probability proportional to the degree of v. This rich-get-richer quality produces the desired power-law behavior in degree distributions.

However, it was recently observed that real networks exhibit two surprising temporal properties [5]: (1) *shrinking diameters*: in many cases, the effective diameter decreases as the network grows; (2) *densification power laws*: networks become denser over time and the average degree increases (the number of edges grows superlinearly in the number of nodes). Moreover, the densification process follows a power-law pattern.

Models like preferential attachment have been assuming otherwise, namely that the average degree is constant over time (i.e. the number of edges grows linearly with the number of nodes) and that the diameter is a slowly growing function of the network size. A primary goal of the recent framework of *stochastic Kronecker graphs* [4] is to capture these two properties while also exhibiting other well-known characteristics like power-law degree distributions. By and large, Kronecker graphs are able to match a wide variety of network proper-

*This work was supervised by Jure Leskovec.

ties; however, there is still scope for improvement, and there are a number of awkwardnesses about the actual model that would be helpful to remove.

The paper is organized as follows. Section 2 defines Kronecker graphs and reviews their properties. Section 3 introduces binomial attribute graphs. Section 4 discusses experiments with real data. Section 5 offers some general perspective on the results and concludes.

2. STOCHASTIC KRONECKER GRAPHS

The construction of stochastic Kronecker graphs is motivated by the observation that many real-world networks follow a fractal-like structure: within each community, there is a subcommunity structure similar to the one observed at the global network level. Defining this recursive property correctly is somewhat subtle, as a number of related methods fail to produce graphs with the necessary properties. The main intuition is to create self-similar graphs recursively by repeatedly taking Kronecker products of some small *initiator matrix* and using this to produce an adjacency matrix for the graph.

Definition 1. Let A be a 2×2 matrix. Then the Kronecker product of A with itself[1] is given by

$$A \otimes A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \otimes \begin{pmatrix} a & b \\ c & d \end{pmatrix} \triangleq \begin{pmatrix} a \cdot A & b \cdot A \\ c \cdot A & d \cdot A \end{pmatrix} = \left(\begin{array}{cc|cc} aa & ab & ba & bb \\ ac & ad & bc & bd \\ \hline ca & cb & da & db \\ cc & cd & dc & dd \end{array}\right),$$

The iterated Kronecker product of a matrix with itself ("Kronecker power") is denoted $A^{[k]}$, so $A^{[2]} = A \otimes A$. The Kronecker product of graphs is defined by taking the Kronecker product of their adjacency matrices. A *Kronecker graph of order k* is defined by $K_1^{[k]}$, where K_1 is called the *Kronecker initiator matrix*.

A number of properties of Kronecker graphs follow immediately from the corresponding properties of the Kronecker product itself, and a number of facts can be proven about them.

A *stochastic Kronecker graph* is obtained in the following manner. Let K_1 be a 2×2 initiator matrix[2] where the entries are real numbers in [0,1]. Form $P_k \triangleq K_1^{[k]}$. Then each entry $p_{uv} \in P_k$ is the probability of an edge between nodes u and v. The matrix P_k can be thought of as specifying a distribution over random (binary) adjacency matrices. To sample from this distribution, sample edge appearance from Bernoulli (p_{uv}) for each pair (u, v). An efficient algorithm exists for estimating the 4 entries of K_1 from data, and a Kronecker graph is considered to perform well if a graph sampled from fitted parameters matches characteristics of the original data with high probability.

[1]The general definition for any matrices A and B generalizes this as one would expect.

[2]Any initiator can be used, but 2×2 suffices in practice, by model selection arguments.

Though this works (perhaps surprisingly) well in practice, there are a few awkward properties:

1. Though the model can be used for forecasting in the sense that one can iterate from P_k to P_{k+1} by taking an additional Kronecker product, this causes the number of "nodes" in the subsequent timestep to double. This is too coarse for many purposes and also conceptually awkward, since identifiability between nodes across time steps is lost. In particular, a single node at time t becomes k nodes at time $t+1$ if the initiator matrix is $k \times k$.

2. Because P_k is formed from an iterated Kronecker product, only graphs with 2^k nodes can be represented. In practice, if a graph has n nodes, one uses the smallest k such that $n < 2^k$, and hopes that the extra nodes do not affect the fit. This usually seems to work in practice, but is not very satisfying.

3. Because of the Kronecker construction, it is not at all obvious how to embellish or modify the model. By contrast, approaches based on probabilistic models are modular and have clear semantics (for example, one can add additional levels to a hierarchical model, change distributional assumptions, and so on).

3. BINOMIAL ATTRIBUTE GRAPHS

The following perspective on the Kronecker construction will help suggest alternatives. Recall that if P_1 is a 2×2, then P_2 is a 4×4, so P_2 nodes can be represented in terms of P_1 nodes by making each P_2 node a *pair* of P_1 nodes: if the P_1 nodes are labeled 0 and 1, then the P_2 nodes are $(0,0), (0,1), (1,0)$, and $(1,1)$. In P_1, the probability of 0 linking to 1 is given by the top-right entry, $P_1(0,1)$. In P_2, what is the probability of $v = (0,0)$ linking to $w = (1,0)$? If v is called the first node, and w the third (think of binary counting), then the probability is $P_2(0,2)$, which is $P_1(0,1) \cdot P_1(0,0)$ by Eqn (1). Now comes the main 'trick': note that $P_1(0,1) \cdot P_1(0,0) = P_1(v_1, w_1) \cdot P_1(v_2, w_2)$; i.e., use the corresponding entries of v and w to *index* into P_1.

Generally, this lets us relate representations of P_k nodes to the correct product of probabilities from P_1. Label nodes of P_k as $v_1, v_2, \ldots, v_{2^k}$, where each node is now a binary k-tuple. Given two nodes v and w in P_k, the probability of an edge can be written

$$p(v, w) = \prod_{i=1}^{k} P_1(v_i, w_i). \quad (1)$$

One can think of the v_i as being binary attributes of v, and P_1 as encoding an attribute similarity matrix that encodes the contribution to the linking probability by nodes agreeing or disagreeing on attributes.

Kronecker graphs implicitly assume that every node has a unique combination of attribute values, and that every combination of attribute values is assigned to some node. In other words, the first node in P_k will always be $(0, 0, \ldots, 0)$ and the final node will be $(1, 1, \ldots, 1)$ and all 2^k possible k-valued binary vectors will be covered in between.

Instead, suppose we do away with the Kronecker product construction entirely, and *sample* binary vectors for each node from some underlying distribution. A coherent probability of an edge can still be obtained using (1) by considering P_1 simply as a parameter matrix rather than an initiator matrix for the Kronecker construction. This suggests the following definition for a particularly simple model of this type.

Definition 2. Let λ be a Bernoulli rate parameter and let V be a fixed set of nodes. Let Θ be a 2×2 parameter matrix, where the entries $\theta_{ij} \in \Theta$ are in $[0, 1]$. Let $K \in \mathbb{N}$.

1. For each $v \in V$, draw K samples from Bernoulli (λ) independently and let A_v be a vector of these samples.
2. For $(u, v) \in V \times V$, determine whether there is an edge $u \to v$ by sampling from

$$u \to v \sim \text{Bernoulli}\left(\prod_{i=1}^{K} \Theta(A_u(i), A_v(i))\right).$$

The result is called a *binomial attribute graph*.

This is one of the simplest possible models one could define that still incorporates a Kronecker-style initiator matrix, in the sense that all edge appearances are completely independent, and the model does not explicit impose any structure at all on the resulting graph. In particular, the recursive Kronecker construction is completely gone, and there is no reason to expect recursive, self-similar structure of the kind produced by the Kronecker graph. Indeed, one would expect this model to perform more like Erdös-Renyi random graphs than something as rigidly structured as a Kronecker graph, but despite this, the results will turn out to be very similar to the Kronecker case.

This construction has the effect that the number of nodes in the model is no longer fixed as 2^k, and the size of attribute vectors K need not be fixed at around $\log_2 |V|$, and that different classes of graphs can be represented, since not all 2^k possible binary vectors need to be accounted for.

Moreover, it is easy to see how to embellish this simplistic model in various ways. For example, we may want the Bernoulli rate λ to vary per node (or by membership in some latent or observed groups), in which case we can assume that each node v is assigned a rate λ_v from some underlying prior distribution, which may be the conjugate prior Beta(α, β) or something more complex. We may not want to restrict some of these values to being binary, in which case we can increase the size of Θ and replace the Bernoulli distributions with multinomials. We may not want edge appearance to be completely independent for all pairs of nodes, in which case the sampling distribution in the second step can be replaced with something more complex that depends on the existing edges in the graph. The attributes need not be considered to be fully latent variables, and binary features from real data could be used. The question in all these cases would be the extent to which these features help fit real data.

4. EXPERIMENTS

The major questions about this model are (1) how similar its graphs are to stochastic Kronecker graphs; (2) how well it is able to match the properties of real data; and (3) how the various parameters and hyperparameters, such as K, affect the quality of the model fit.

For background, Kronecker graphs have been assessed in the following manner: first, fit the parameters to a particular dataset; second, generate a synthetic graph using those parameters; third, compare the generated graph and the real graph in various ways to assess the fit. Some of the metrics used are the following:

1. *Degree distributions*: one can compare the in-degree and out-degree distributions (which generally follow power laws).
2. *Small diameter*: is the diameter of the synthetic graph close to the true diameter? Since the diameter is susceptible to outliers, a modified statistic called the *effective diameter* [4] is used instead.
3. *Hop plots*: these extend the notion of diameter by plotting the number of reachable pairs within h hops as a function of the number of hops h. This gives a sense of how quickly node neighborhoods expand with the number of hops.
4. *Singular values*: this is a plot of the singular values of the adjacency matrix, which has also been found to have consistent structure across datasets.
5. *Node triangle participation*: a local statistic measuring how many triangles each node participates in; we can plot the statistic by node degree.
6. *Clustering coefficient*: a long-used measure of local clustering behavior [7]; we can plot the average clustering coefficient by node degree.

These various plots can be examined visually for qualitative similarity (which is often enough to outperform many other models, which are completely hopeless for some of these properties), or one can use a goodness-of-fit statistic like the Kolmogorov-Smirnov D-statistic for a quantitative measure of discrepancy between the synthetic and true distributions of the statistics above. One could also use general measures of discrepancy between distributions, such as the KL-divergence or Hellinger distance.

To make binomial attribute graphs as directly comparable to Kronecker graphs as possible, we set Θ to the fitted[3] values of the Kronecker initiator matrix P_1, set $K = \lceil \log_2 V \rceil$ (so the A_v are the same size as they would be in the Kronecker case), and then fit λ so the expected density of the binomial attribute graph matches the true one. This is done by minimizing

$$((V^2 - V) \cdot (\theta_{00} \cdot (1-\lambda)^2 + (\theta_{01} + \theta_{10}) \cdot \lambda \cdot (1-\lambda) + \theta_{11} \cdot \lambda^2)^K - E)^2$$

[3]Kronecker graph parameters are fit via KronFit [4], a Metropolis-Hastings algorithm that exploits the recursion.

with respect to λ, subject to $\lambda \in [0, 1]$. Here V and E are the number of nodes and edges in the real data. This problem can be solved quickly using nonlinear bound-constrained optimization algorithms such as L-BFGS-B [2].

4.1 Autonomous Systems

Here, we consider a static dataset consisting of connectivity among internet Autonomous Systems [6]. The internet's router graph can be organized into subgraphs called Autonomous Systems, and a communication network among these can be formed by examining BGP logs that show which pairs exchange traffic. This data has 6474 nodes and 26467 edges. See Figure 1 for sample fits. These graphs reveal that both models are substantially better than Erdös-Renyi graphs (which is expected for Kronecker graphs, but not necessarily binomial graphs), and are extremely similar to each other across all the various network statistics. The diameters of the graphs are also very similar (between 8 and 11 in all cases).

Though in this particular case all the statistics match quite well, this is not necessarily consistent across datasets. In particular, these models can do poorly on clustering and triad participation, as those statistics measure very local properties, and these models are highly 'global' in flavor. Poorer results of this kind are seen on datasets like citation and collaboration networks from arXiv (results not shown for space reasons).

4.2 Sparsity Patterns

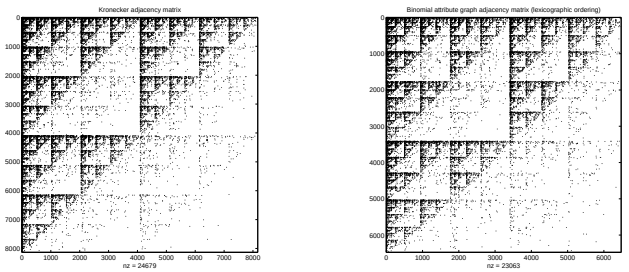
Because of the nature of the Kronecker product, the adjacency matrix of a stochastic Kronecker graph will have a distinctive, fractal-like sparsity pattern that can be seen via Matlab `spy` plots. To directly compare binomial and Kronecker graphs to each other, we can ask whether the graphs produced by the binomial model have any similar quality.

Because there is no explicit node ordering in the binomial model, unlike the Kronecker model, the matrix will look like random noise, but by *lexicographically reordering* the nodes in the binomial graph by their attribute vectors A_v (treat them like binary numbers and sort in ascending order going left to right), the adjacency matrix can be compared to the Kronecker case.

Figure 2 shows that despite not explicitly modeling any fractal structure, the binomial model can recover this structure by virtue of having a "wide enough variety" of per-node attribute vectors and the same parameter/initiator matrix. This provides some intuition as to why this rather unstructured model is so similar to one as rigid as Kronecker graphs.

4.3 Varying Model Complexity

Is the quality of the model fit affected by changing K to be larger or smaller than $\lceil \log_2 V \rceil$? To examine this, we consider the Autonomous Systems data, which normally uses $K = 13$, and considered values of K between 7 and 17. For each of these different values of K, the parameter λ is refit. The main result, as seen in Figure 3, is that varying K does not make a substantial difference to the quality of the model fit. Table 1 shows Kolmogorov-Smirnov statistics that compare each of the listed distributions against the distributions from the Autonomous Systems data.



(a) Kronecker graph (b) Binomial graph

Figure 2: Sparsity patterns in adjacency matrices.

5. CONCLUSIONS

We considered many experiments in addition to these, and this subset has been selected just to give some flavor of the study. To some degree, the fact that binomial attribute graphs work at all raises more questions than it answers. One immediately useful conclusion, however, is that the reason Kronecker graphs work is *not* because of the actual Kronecker construction – this construction actually turns out to be a red herring. All the modeling power, in some sense, is contained in the Θ matrix, which encodes a high-level summary of the structure of the network. For example, if $\Theta = [0.9, 0.5; 0.5, 0.1]$, which is a common template, this suggests that the graph roughly consists of a tightly linked core (top left) with an intermediate number of cross-links (off-diagonals) with a sparsely connected periphery (bottom right). In other words, the fitted Θ itself encodes some structure of the network, and this is the primary driver behind the modeling power of Kronecker graphs. It is rather surprising that these four parameters do so much. Binomial graphs work similarly well because the Bernoulli (λ) sampling scheme produces a similar variety of attribute vectors as the Kronecker graph, so as long as the distribution of attribute vectors has sufficient variance, this is enough to allow the Θ matrix to impose its structure (Figure 2 being one way of seeing this).

Future work in this direction could take two directions: first, one could attempt to design a more sophisticated variant of the binomial attribute model that performs better at some tasks; second, it would be worth trying to understand at a deeper level why the binomial model works as well as it does. In particular, a better theoretical understanding of why this scheme suffices may help in designing improved generative network models.

6. REFERENCES

[1] A.-L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 1999.

[2] R. H. Byrd, P. Lu, and J. Nocedal. A limited memory algorithm for bound constrained optimization. In *SIAM Journal on Scientific and Statistical Computing*, 1995.

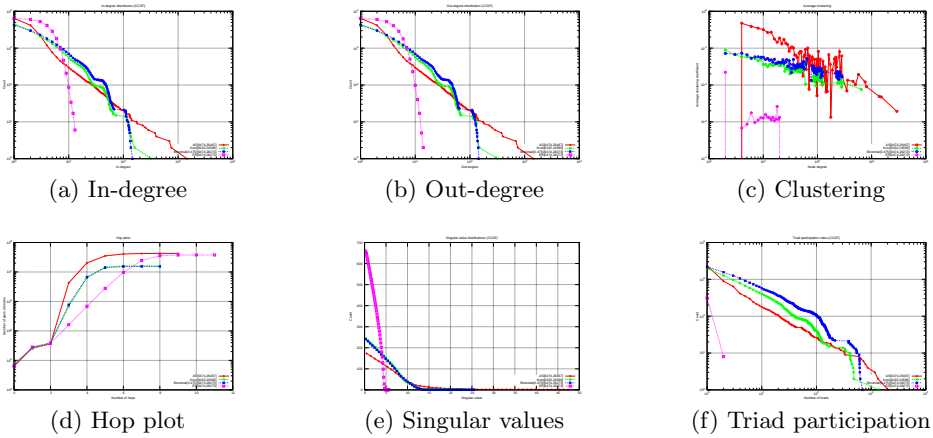
[3] P. Erdös and A. Renyi. On the evolution of random graphs. In *Publication of the Mathematical Institute of the Hungarian Academy of Science*, 1960.

[4] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani. Kronecker graphs: An approach to modeling networks. arXiv:0812.4905v2, 2009.

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

[6] U. of Oregon. Routeviews project.

[7] D. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 1998.



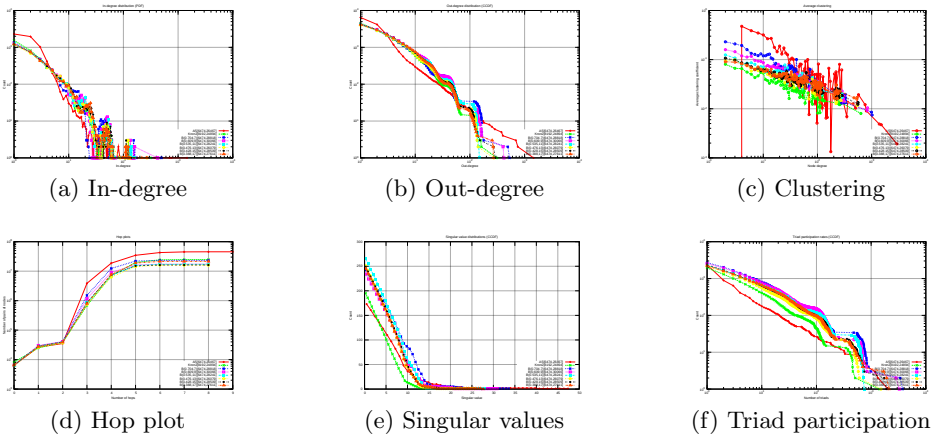
(a) In-degree (b) Out-degree (c) Clustering

(d) Hop plot (e) Singular values (f) Triad participation

Figure 1: Fits for real (red), Kronecker (green), binomial (blue), and Erdös-Renyi (pink) graphs.

	In	Out	Cluster	Singular	Triad
Kron[8192,24894]	0.449341	0.448364	0.124876	0.212924	0.130312
B(0.704,7)[6474,28818]	0.342910	0.351251	**0.080156**	0.176205	0.176861
B(0.609,9)[6474,30066]	**0.342138**	0.345845	0.095077	0.115706	0.158171
B(0.535,11)[6474,28241]	0.365153	0.354032	0.127567	0.130091	0.118628
B(0.476,13)[6474,26079]	0.381372	0.381989	0.144830	0.089906	**0.098085**
B(0.428,15)[6474,28509]	0.349707	**0.343528**	0.147132	**0.073682**	0.123108
B(0.388,17)[6474,27611]	0.351715	0.356348	0.175157	0.081232	0.111677

Table 1: Kolmogorov-Smirnov statistics comparing fitted distributions to real data. The numbers in parentheses are parameters λ and K and the numbers in brackets are node and edge counts.



(a) In-degree (b) Out-degree (c) Clustering

(d) Hop plot (e) Singular values (f) Triad participation

Figure 3: Fits for real (red), Kronecker (green), and binomial graphs with $K \in \{7, 9, 11, 13, 15, 17\}$.