Rigorous Analysis of Kronecker Graphs and their Algorithms

Bharath Ramsundar

October 23, 2012

1 Reaction Paper

Real world graphs have been observed to display a number of surprising properties. These properties include heavy-tails for in- and out-degree distributions, small diameters, and a densification law [5]. These features do not arise from the classical Erdos-Renyi random graph model [1]. To address these difficulties, Kronecker Graphs were first introduced in [5] as a new method of generating graphs which match network data. In this literature survey, we review two works on Kronecker Graphs [8], [6] which respectively study the theoretical properties of Kronecker Graphs and introduce new learning algorithms to fit Kronecker Graphs to real Network Data. We also review the classical work by Erdos and Renyi [1] to gain inspiration from the techniques of the old masters.

The first paper we consider is *Kronecker Graphs: An Approach to modeling networks* [6]. This paper reviews the use of Kronecker graphs as useful tools for modeling networks (the results covered were first introduced in earlier works such as [5]). Kronecker graphs provide methods of "multiplying" two graphs together via the tensor (kronecker) product on the adjacency matrix. The paper also covers the formalism of Stochastic Kronecker graphs. Such graphs are constructed not from adjacency matrices, but rather from kronecker products of probability matrices, where the (i, j)-th entry is the probability of an edge between i and j existing. The probability matrix associated with a Kronecker graph is also called its initiator matrix. Note that each such probability matrix defines a distribution over graphs and not a single graph. A fast approximate algorithm for generating Kronecker Graphs is presented. The theoretical guarantees of this algorithm are not studied.

This work also considers the problem of modeling real-world network data with Kronecker graphs. It presents a MLE algorithm **KronFit** that fits an initiator matrix to match a Kronecker graph to a real world network. An interesting aspect of this algorithm is that it deals inference upon an enormous, combinatorial space (the generated Kronecker graph). However, by clever exploitation of the recursive structure of the graph generation, efficient algorithms are achieved. One interesting challenge is that the likelihood in this space cannot be evaluated explicitly (for doing so would require a summation over an exponentially large set), but is rather evaluated through clever use of Markov Chain Monte Carlo methods and approximate gradient descent.

A Markov chain is used to sample random permutations which can permute the elements of a proposed Kronecker Graph. These permutations are necessary since graphs which only differ in the labelling of their vertices should not be considered as different. The convergence of this chain is mainly evaluated through empirical methods, so some interesting work might involve studying the mixing time of this Markov Chain using more analytical methods. The evaluation of likelihood functions is sped up through use of a Taylor Approximation. Explicit error bounds are not presented for this construction. The size of the initiator matrix for the fitted Kronecker graph is chosen through use of the Bayesian Information Criterion.

These algorithms are verified experimentally through use of synthetic and real data. The synthetic data is generated from hand-picked initiator matrices. The paper claims that the picked initiator matrices match those calculated from real-world data. Convergence of the Markov Chain is evaluated by plotting the autocorrelation function and another measure called scale reduction. The convergence of the log-likelihood sampling scheme is evaluated experimentally as well. A simple argument shows that the maximum likelihood problem at hand is nonconvex. However, the nonconvexity of the space is noted to have a nice structure, for the multiple global minima are in some way all the same minima (up to permutation). Experiments show that 98% of runs converged to the true matrix. Experiments are also performed which compare the properties of the Kronecker graphs to real data-sets such as citation networks and autonomous internet systems data. From a first order, such fits appear interesting, since the desired heavy tail distributions for degrees and densification are experimentally visible.

At a high level, this summary paper provides empirical and anecdotal evidence for why Kronecker graphs provide useful modelling tools. However, only the most elementary properties of such graphs are analyzed mathematically. More rigorous mathematical analysis was undertaken in the paper *Stochastic Kronecker Graphs* [8]. This work performs an analysis of the properties of Stochastic Kronecker graphs. However, it restricts itself to the case where the initiator matrices are 2×2. Various necessary and sufficient conditions are given for the existence of a giant connected component. This work proves that decentralized search algorithms cannot efficiently find paths through Stochastic Kronecker graphs. This condition raises an interesting point. Can a modification of the Stochastic Kronecker Graph which is searchable be found?

The main technical lemma of the work is a condition on the existence of a giant connected component in Kronecker graphs. This lemma is proved by exploiting facts about min-cuts and states that if the min-cut size of the weighted graph defined by an adjacency matrix P is at least $c \ln n$, then a Kronecker graph generated according to P will be connected with high probability. This proof follows from an application of the Chernoff bound to a fact about the number of min-cuts in a graph. Some further facts about connectivity are developed. These facts are used, along with some knowledge of the diameter of Erdos-Renyi random graphs, in order to prove that the diamater of Kronecker graphs will be constant with high probability.

This work makes a number of seemingly restrictive assumptions upon the initiator matrix P. The authors argue that their restrictions are in line with empirical results found in earlier works. It is not clear to me how realistic their claims are. Would it be possible to extend their results to the case of arbitrary initiator matrices? Would doing so be theoretically interesting? Another idea is that the proofs offered in [8] are quite messy and rely on repeated applications of Chernoff bounds. Could a more geometric argument provide better intuition?

The last work we survey is a classic from the mathematical history of random graph theory. *On Random Graphs* [1] by Erdos and Renyi provides basic analysis of the $G(n, m)$ model. One interesting fact that this work proves is that if the number of edges in the graph is N_c where

$$N_c = \frac{1}{2} n \ln n + cn$$

then the number of points outside the great component is distributed as a Poisson distribution with mean e^{-2c}. This paper also defines the following random process: at each step, add an edge with equal probability to any open pair of vertices; continue till the graph is connected. Then the number of edges required to terminate the process (i.e. connect the graph) is clustered closely around $n \ln n$ once again. This dependence on $n \ln n$ might be the phase transition associated with the Erdos-Renyi model. The paper proves these facts by using the following lemma

Lemma 1.1. *As the number of vertices n goes to infinity, with probability* 1*, a random graph generated from $G(n, N_c)$ has a giant component with $n - k$ nodes and k disconnected individual components.*

This lemma is very interesting since it shows that the idea of a giant component was well known to the mathematical community almost 60 years before now. This work is also interesting for the form of the mathematical arguments used. Direct proofs without recourse to recursive formulae are used to prove this result.

2 Project Proposal

The problem we will be solving in this project is that of analyzing the algorithms used to fit Kronecker Graphs to real network data. These algorithms provide interesting examples of Machine learning models which learn from very high dimensional, combinatorial data. The work in [6] attacks this problem from an empirical approach, through a variety of simulations and empirical measures. Our work in this project will be to complement this experimental approach with a rigorous theoretical analysis.

To start this analysis, we first plan to implement the **KronFit** algorithm and test it on synthetic and real datasets mentioned in [6]. These datasets include citation and autonomous systems data and are available online at [3]. To develop empirical intuition about the performance, we will graphically analyze the autocorrelation and the scale reduction for these algorithms. We will also verify various configurations of parameters claimed to be optimal in [6]. The goal of this exercise is to gain an intuitive understanding of the Markov chains and Maximum Likelihood approximations used by **KronFit**. The projected completion of this implementation part of the project would be at the progress report.

The next task is to mathematically analyze the convergence and runtime of these algorithms. There are a number of known analytic techniques to calculate the mixing times of Markov Chains [7]. The first step in the theoretical analysis would be to conduct a thorough study of whether each of these techniques would apply to the Markov Chains required for **KronFit**. If none of these techniques leads to a theoretical analysis, we will take steps towards developing the techniques required for our analysis.

The second task is to understand the nature of the likelihood approximations used to compute the gradient efficiently in **KronFit**. These sorts of approximate gradient methods are becoming more prevalent, for machine learning methods are now applied to ever more complex models most of which have no simple analytic descriptions [4]. A few techniques have been developed to bound these approximations. We will explore whether these techniques suffice to analyze **KronFit**.

The final theoretical analysis we will undertake is a more rigorous investigation of the mathematical structure of Stochastic Kronecker Graphs. Prior work in [8] performs analysis of Stochastic Kronecker graphs, but only for the limited case of 2×2 initiator matrices. We will consider the question of whether the proofs can be generalized to arbitrary initiator graphs.

Since the second part of this project is theoretical in nature, the exact results cannot be guaranteed. Proofs may turn out to be significantly more difficult to obtain than originally envisioned. We will however guarantee thorough literature surveys of known results for each of the previous subtasks. Specifically, we will perform a survey of known MCMC mixing bound techniques as they relate to the combinatorial spaces explored by the MCMC step in **KronFit**. For the likelihood approximation steps, we can discuss known techniques from the Sequential Monte Carlo Literature [4] and how these techniques may be applied to the Graph-Isomorphism type problems found in

KronFit. Of the topics listed above, we are least confident of achieving a good proof of properties of Stochastic Kronecker Products for arbitrary initiator graphs. It is likely that the prior work [8] ran into serious technical difficulties that obstructed them from general proofs. However, as mentioned before, if we fail to achieve the desired proofs, we will provide a discussion of why well known matrix techniques such as spectral analysis [7] or various decompositions don't help here.

We now consider some other analyses which could be undertaken if time permits. One interesting possible task would be considering how to generalize the parameter learning algorithms of **Kronfit** to other learning settings. There are a number of tasks where learning parameters from large combinatorial spaces and more general algorithms could lead to interesting results. Yet another task would involve extending the precise analysis performed by Erdos and Renyi (specifically the explicit formula implying a poisson distribution for the number of connected components) to the Kronkecker graph model. Other interesting work might involve extending other analyses performed by Erdos and Renyi such [2].

References

[1] P. Erdos and A. Renyi. On random graphs i. 1959.

[2] P. Erdos and A. Renyi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutat Int. Közl*, 1960.

[3] J. Leskovec et al. http://snap.stanford.edu/data/index.html.

[4] N. Kantas. An overview of sequential monte carlo methods for parameter estimation in general state-space models. In *Proc. IFAC Symposium on System Identification (SYSID)*, 2009.

[5] J. Leskovec, D. Chakrabarti, J. Kleinberg, and C. Faloutsos. Realistic, mathematically tractable graph generation and evolution, using kronecker multiplications. In *Knowledge Discovery in Database (PKDD)*, 2005.

[6] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani. Kronecker graphs: An approach to modeling networks. *Journal of Machine Learning Research (JMLR)*, 2010.

[7] D. Levin, Y. Peres, and E. Wilmer. *Markov Chains and Mixing Times*. Americal Mathematical Society, 2008.

[8] M. Mahdian and Y. Xu. Stochastic kronecker graphs. In *Fifth Workshop on Algorithms and Models for the Web Graph (WAW)*, 2007.