RANDOM DOT PRODUCT GRAPHS: A FLEXIBLE MODEL FOR COMPLEX NETWORKS

A Thesis
Presented to
The Academic Faculty

by

Stephen J. Young

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in
Algorithms, Combinatorics, and Optimization

School of Mathematics
Georgia Institute of Technology
December 2008

RANDOM DOT PRODUCT GRAPHS: A FLEXIBLE MODEL FOR COMPLEX NETWORKS

Approved by:

Professor Milena Mihail, Advisor
College of Computing
Georgia Institute of Technology

Professor Prasad Tetali
School of Mathematics and College of Computing
Georgia Institute of Technology

Professor William T. Trotter
School of Mathematics
Georgia Institute of Technology

Professor Xingxing Yu
School of Mathematics
Georgia Institute of Technology

Professor Joel Sokol
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Reader: Professor Linyuan Lu
Department of Mathematics
University of South Carolina

Date Approved: 14 November 2008

For her patience, understanding, and love

to Naomi.

ACKNOWLEDGEMENTS

I would first of all like to thank the School of Mathematics and the ACO program for providing me this opportunity. I would especially like to thank Prof. Luca Dieci for his support and understanding when things were not going quite as planned. Also, I would like to thank the staff of the School of Mathematics for their phenomenal support and their ability of making administrative headaches disappear.

I would like to thank my committee for the time and effort and for all they have taught me directly and indirectly. I would especially like to thank Prof. Linyuan Lu for his thoughtful remarks regarding this thesis and for taking the time to serve as reader.

I would like to thank all of my friends in the ACO program and the School of Mathematics, especially Mark, Mitch, Csaba, Dave, and Noah. Thank you guys for listening to me spout my ideas and helping refine them.

I also thank my wife and family for trying to understand what I do and why I do it, and encouraging me to see it through to the end. I also thank the people too numerous to name whose constant belief in me has kept me going throughout this process.

I especially want to thank Prof. Tom Trotter for accepting me as a "half"-student. I am deeply thankful for all you have taught me by word and example, both as a mathematician and as a person.

Most of all I would like to thank my advisor, Milena Mihail, for accepting someone she barely knew with an out of the blue idea for a thesis topic, for having the patience and perserverence to see me through the process even when it seemed to be going nowhere, and for having the insight to know what I needed, even when I didn't know myself. Thank you.

SUMMARY

Over the last twenty years, as biological, technological, and social networks have risen in prominence and importance, the study of complex networks has attracted researchers from a wide range of fields. As a result, there is a large and diverse body of literature concerning the properties and development of models for complex networks. However, many of the models that have been previously developed, although quite successful at capturing many observed properties of complex networks, have failed to capture the fundamental semantics of the networks. In this thesis, we propose a robust and general model for complex networks that incorporates at a fundamental level semantic information. We show that for a large range of average degrees and with a suitable choice of parameters, this model exhibits the three hallmark properties of complex networks: small diameter, clustering, and skewed degree distribution. Additionally, we provide a structural interpretation of assortativity and apply this strucutral assortativity to the random dot product graph model. We also extend the results of Chung, Lu, and Vu on the spectral gap of the expected degree sequence model to a general class of random graph models with independent edges. We apply this result to the recently developed Stochastic Kronecker graph model of Leskovec, Chakrabarti, Kleinberg, and Faloutsos.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

At its heart, the study of complex networks is the quest for identifying and modeling the presence of structure where there is no a priori predictive or explanatory reason for such structure. This phenomena occurs in a wide range of contexts, from "optimized" networks such as the structure of the physical layer of the Internet and the power grid, to evolved networks such as gene-protein interaction networks and food webs, and from networks with "costly" edges such as sexual contact networks and collaboration networks, to networks with "free" edges such as the LiveJournal friend network and the World Wide Web. For a survey of such results see [6, 16, 20, 24, 50, 33]. Given that networks with such dissimilar sources exhibit similar behaviors, such as power-law degree distribution, clustering, and small diameter, it is natural to hope for a model of the various complex networks that would have explanatory power for all such networks. In many ways the Holy Grail of the study of complex networks is a flexible and robust model that, by varying a small number of well understood parameters, can encapsulates a large class (hopefully all) of complex networks. Further, such a model should respect the semantic content of the underlying network in order that the parameters of the "best"-fit model retain the semantic information of the network. The goal of this thesis is to provide a robust and flexible model that will hopefully be a stepping stone towards the eventual discovery of such a "Holy Grail" model.

1.1 Previous Work

In order to provide context for the development of the random dot product graph model we briefly survey some of the results on a selection of the earlier and more well known models for complex networks. By no means do we wish to imply that this list is comprehensive, rather we feel that these models give a flavor for the types of random graphs that have been developed previously in an to attempt to understand and model complex networks. For a more comprehensive survey see [16, 20, 24, 33].

Preferential Attachment The preferential attachment model was one of the first models for complex networks to gain prominence. It was proposed as a model for complex networks by Barabási and Albert [7] who describe the random graph as the result of sequentially adding vertices and distributing the edges incident to the new vertex proportionally to the degree of the existing vertices. The work of Bollobás and Riordan [14] was the first of many works to formalize this preferential attachment idea. They formalize the preferential attachment model as deriving from randomized linearized chord diagrams and use this formalization to show that the diameter is $\Theta\left(\frac{\log(n)}{\log(\log(n))}\right)$. In [46] Mihail, Papadimitriou, and Saberi are able to show that the preferential attachment model exhibits constant conductance and spectral gap and hence the network has good congestion and mixing properties. These results, combined with the natural description, would seem to indicate that the preferential attachment model may be a good fit for many complex networks. However, in the first chapter of [16] Bollobás and Riordan prove that the clustering coefficient of the graph resulting from the linearized chord diagram formalization of the preferential attachment model tends towards zero with the number of vertices. Furthermore, Bollobás, Riordan, Spencer, and Tusnády show that the degree sequence follows a power law with exponent precisely three [15]. Thus, although the preferential attachment model is semantically appealing, it is hard to argue that it accurately models many complex networks with its fixed degree distribution and asymptotic lack of clustering.

Geometric Preferential Attachment In an attempt to address the existance of small separators in many complex networks [43, 44], Flaxman, Frieze, and Vera developed a modification of the preferential attachment model incorporating an underlying geometry. Specifically, vertices are sequentially and randomly distributed on the sphere of radius $\frac{1}{2\sqrt{\pi}}$ and a new vertex is randomly assigned a set of m neighbors where each vertex gains a new neighbor with probability proportional to its degree and a function of the distance to the new vertex [28, 29]. They are able to show that the resulting model exhibits a power-law degree distribution and is connected with small diameter. They are also able to show that there are small separators in the resulting graph with high probability, however these separators are inherently geometric (the great circle) and split the graph roughly uniformly. This is contrast to the the variety of observed behavior in many complex networks, see [25, 43, 44].

Copying The copying model is an attempt to model the biological processes underlying gene replication or the natural evolution of the Internet, as an explanation for the structure of complex networks [21, 37, 38]. The basic idea behind the copying model is, that as a new node enters the network it selects according to some distribution, potentially dependent on the current graph, a vertex v to copy and then it chooses its neighborhood as the neighborhood of v up to a mutation factor that can add and delete vertices from the neighborhood set. In [21, 37, 38] it was shown that with partial duplication the copying model exhibits a power-law degree distribution.

Configurational The configurational random graph model on a degree sequence $d_1 \leq d_2 \leq \cdots \leq d_n$ is formed by randomly choosing a perfect matching on

$$v_{1,1}, v_{1,2}, \cdots, v_{1,d_1}, v_{2,1}, v_{2,2}, \cdots, v_{2,d_2}, \cdots, v_{n,d_n}.$$

Given this perfect matching, there is an edge between vertices s and t if there are some i and j such that $\{v_{s,i}, v_{t,j}\}$ is an edge in the perfect matching. Note that the underlying multigraph for this graph will have the given degree sequence, but the graph will likely not have the desired degree sequence. In [47, 48] it is shown that there is a thresholding function for this model which delineates the degree sequences for which there is a giant component. Gkantsidis, Mihail, and Saberi were able to show that if $\sum_i d_i$ is $\mathcal{O}(n)$, $\max_i d_i$ is $\mathcal{O}(\sqrt{n})$, and if the minimum degree is at least 3, then the resulting graph has good conductance and it is possible to route $\mathcal{O}(d_i d_j)$ flow between every pair of vertices i and j with near optimal congestion [32].

In all of the models that have been mentioned thus far, we note that presence of many of the edges are clearly dependent on the arrangement of other edges. It is worth noting that all of these models with dependent edges can be efficiently generated, either by sequential generation (preferential attachment, geometric preferential attachment, and copying) or by cleverly exploiting the dependencies between edges (configurational). This is in contrast to the following models which have edges that are independent of the other edges in the graph and will require $\Theta\left(n^2\right)$ random samples to generate as a result.

Expected Degree Sequence Given a desired degree distribution $1 \leq d_1 \leq d_2 \leq \cdots \leq d_{n-1} \leq d_n \leq \sqrt{n}$ the expected degree sequence model has each edge $\{i, j\}$ present independently with probability $\frac{d_i d_j}{\sum_k d_k}$, and thus, allowing self-loops, the expected degree of vertex i is d_i. It is clear that if the original degree sequence is a power-law then the expected degree-sequence is also a power-law, however, the concentration results have not been established. In [18, 19] Chung and Lu analyze the emergence of the giant component within this model and show that if the exponent of the power-law is within the standard range of 2 to 3, then the average distance between pairs of vertices is $\log(\log(n))$ while the diameter of the graph is $\log(n)$. In addition, together with Vu, they show that, with certain assumptions about the degree sequence, the expected degree sequence model generates graphs with good conductance and spectral gap, and hence behaves well algorithmically [22]. However, a simple calculation shows that this graph does not exhibit any clustering, irregardless of the choice of degree sequence.

Stochastic Kronecker The Stochastic Kronecker graph was recently proposed as a model for complex networks by Leskovec, Chakrabarti, Kleinberg, and Faloutsos [40]. In the Stochastic Kronecker framework the edge from i to j is present independently with probability equal to p_{ij} and the matrix P is formed by repeated Kronecker multiplication of some generating matrix. For instance if the generating matrix is

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

then the Kronecker product with P would be

$$\begin{bmatrix} aP & bP \\ cP & dP \end{bmatrix}.$$

Even though this results in an inherently multinomial degree distribution, Leskovec and Faloutsos are able to show that using,

$$\begin{bmatrix} .98 & .58 \\ .58 & .06 \end{bmatrix}$$

as a generating matrix for a Stochastic Kronecker graph yields a graph that fits the Internet at the autonomous system level fairly well [41]. In addition, the Stochastic Kronecker model exhibits "densification" as observed in real world networks by Leskovec, Kleinberg, and Faloutsos [42]. In [45], Mahdain and Yu use a general result about the connectivity of random graphs with independent edges to determine the conditions for which a Stochastic Kronecker graph generated by a 2×2 symmetric matrix is connected. Additionally, they show that, under certain conditions a Stochastic Kronecker graph has small diameter. Further, they show that there is some constant c such that no localized routing scheme can find paths of length less than n^c, where n is the number of vertices.

Inhomogeneous Random Graphs Bollobás, Jansen, and Riordan recently, and independent of this work, introduced a remarkably elegant and general sparse random graph model that attempts to capture the flexibility and robustness required to model complex networks [13]. In essence, each edge $\{i, j\}$ is assigned a weight w_{ij} and each edge $\{i, j\}$ is present independently with probability $\min\left\{\frac{w_{ij}}{n}, 1\right\}$. Similarly to the random dot product graph model, the weights w_{ij} are chosen randomly but not independently, reflecting some underlying semantic content present at each vertex. They incorporate additional semantics via a "kernel" function on the base space. We will compare this model with the random dot product graph model in more depth in Section 2.4.

1.2 Contributions of this Thesis

In this thesis we attempt to develop a model for complex networks that is flexible, robust, and can be meaningfully fitted to a large class of complex networks. Since one of the goals of this thesis is to develop a model that can be fit to various complex networks it is natural to fallback on the methodologies used for large scale data analysis. There is a long history of using linear algebraic techniques to assign vectors in some high-dimensional space to entities with the inner product as the measure of distance [5, 52, 51]. Thus it is natural to allow the inner product to be the fundamental object of a random graph model that is attempting to capture semantic information. Combining this with a desire for an extremely flexible model, we arrive naturally at the following informal construction. Assign to each vertex a vector in $\mathbb{R}^d$ from a general distribution and then allow each edge to be present independently with probability proportional to the inner product of the end points of the edge. We refer to such a graph as a random dot product graph, and we will formally define such graphs in Chapter 2. We note that the parameter d is fixed and that there is a natural interpretation of the vector associated to a vertex as representing the vertex' "interest" or "properties". Thus two vectors with similar interests or properties would be more likely to be connected by an edge. From an applications point of view the restriction to a fixed finite dimension for the vectors follows naturally from the finiteness of computational resources. Technically, we will see that the fact that the dimension is fixed and finite is a key, but subtle, requirement for the proof of the diameter results (Theorem 2.1 and Theorem 3.2).

At this point, it is worth observing that it is clear that the random dot product graph model is a generalization of the standard Erdős-Rényí random graph. In addition, it is in some limited sense of a generalization of models such as the expected degree sequence model [22] and the model developed by Caldarelli, et al. using vertex intrinsic fitness [17]. It also fits into the general framework for random graphs with independent edges introduced independently by Sönderberg [54] and Kraetzl, Nickel, and Scheinerman [36]. The recent work by Bollobás, et al. on inhomogeneous random graphs can also be phrased naturally within this general framework [13]. However, as we will see in Section 2.4, the random dot product model and the inhomogeneous graph model will cover disjoint aspects of the fully general model.

We generalize, in Chapter 2, the definitions of random dot product graphs first introduced by Kraetzl, Nickel, and Scheinerman [36] to include a large class of distributions on $\mathbb{R}^d$ and general sparsity. In their work Kraetzl, et al. deal formally only with average degree $\Theta(n)$ and with one-dimensional vectors distributed according to $\mathcal{U}^\alpha(0,1)$, where $\mathbb{P}(\mathcal{U}^\alpha(0,1) \leq t) = t^{\frac{1}{\alpha}}$ for $t \in [0,1]$. Within this context they were able to show that there is a $1 - o(1)$ fraction of the graph which has diameter at most two with high probability. Further, they show that the graph exhibits positive clustering, specifically

$$\left(\frac{\alpha+1}{2\alpha+1}\right)^2 = \mathbb{P}(u \sim w \mid u \sim v, v \sim w) > \mathbb{P}(u \sim w) = \frac{1}{(\alpha+1)^2}.$$

Finally, they show that for $k \geq n^{\frac{23}{24}}$ the expected number of vertices v with $(1 - n^{-\frac{1}{12}})k \leq$

$\deg(v) \leq (1 + n^{-\frac{1}{12}})k$ is equal to

$$\frac{n}{\alpha}\left(\frac{\alpha+1}{\alpha}\right)^{\frac{1}{\alpha}} 2k\mathcal{O}\left(n^{-\frac{1}{12}}\right) k^{\frac{1}{\alpha}-1},$$

with probability approaching 1 as $n \to \infty$. In particular, they show that the "log-histogram" of the degree distribution has heavy-tails and those tails have a power-law distribution.

We will extend the results and the ideas contained within the work of Kraetzl et al. to general distributions μ on $\mathbb{R}^d$ satisfying a few natural conditions and average degree $\omega(1)$. In particular, we created a general reduction scheme for the diameter of random dot product graphs to Erdős-Rényí graphs with similar average degree by extending the ideas implicitly present in [36] and explicitly expressed by Young and Scheinerman in [59, 60]. With respect to clustering, we are able to prove positive clustering for non-constant distributions with arbitrary average degree by extending the proof that appeared in [59, 60]. Finally, if μ satisfies certain mild conditions, we are able to replace the previously best known formula for the degree distribution,

$$\mathbb{P}(\deg(v) = k) = \int \binom{n-1}{k} \left(\frac{\langle \mathbb{E}[\mu], X\rangle}{g(n)}\right)^k \left(1 - \frac{\langle \mathbb{E}[\mu], X\rangle}{g(n)}\right)^{n-1-k} d\mu(X), \tag{1.1}$$

with

$$\mathbb{P}(|\deg(v) - k| \leq \delta k) = \mu^{\langle\cdot,\cdot\rangle}\left(\left[\frac{g(n)(1-\delta)k}{n-1}, \frac{g(n)(1+\delta)k}{n-1}\right]\right) + o(1),$$

where $\mu^{\langle\cdot,\cdot\rangle}$ is a probability measure on $[0,1]$ derived from μ. Furthermore, we are able to show that if k is $\omega(1)$ then the error term can be refined to $\mathcal{O}\left(\sqrt{\frac{\ln(k)}{k}}\right)$. To our knowledge this makes the full generality of random dot product graphs on appropriately chosen μ one of the first models to have all three of the hallmark properties of complex networks; small diameter, clustering, and skewed degree distributions.

Chapter 3 deals with the natural directed generalization of random dot product graphs, which first appeared in [60]. This work shows that for directed random dot product graphs with average in and out degree $\Theta(n)$ and mild conditions on the distributions, there is an arbitrarily large fraction of the graph that is strongly connected with diameter 5, that in many ways the directed random dot product graph exhibits clustering, and that the in and out degree can be expressed in terms of an integral similar to Equation (1.1). In

Section 3.1 we extend these results to show that allowing arbitrary average degree does not affect the clustering of the random dot product graph. Section 3.2 generalizes the ideas of the diameter result of [60] and Theorem 2.1, to again provide a reduction to the directed diameter of a directed Erdős-Rényí random graph. Finally in Section 3.3 we simplify the formula for the in-degree and the out-degree of a vertex from

$$\mathbb{P}\big(\deg^+(v) = k\big) = \int \binom{n-1}{k} \left(\frac{\langle \mathbb{E}[\mu], Y\rangle}{g(n)}\right)^k \left(1 - \frac{\langle \mathbb{E}[\mu], Y\rangle}{g(n)}\right)^{n-1-k} d\nu(Y) \quad \text{and}$$
$$\mathbb{P}\big(\deg^-(v) = k\big) = \int \binom{n-1}{k} \left(\frac{\langle X, \mathbb{E}[\nu]\rangle}{g(n)}\right)^k \left(1 - \frac{\langle X, \mathbb{E}[\nu]\rangle}{g(n)}\right)^{n-1-k} d\mu(X)$$

to

$$\mathbb{P}\big(\big|\deg^+(v) - k\big| \le \delta k\big) = \nu^{\langle \cdot, \mathbb{E}[\mu]\rangle}\left(\left[\frac{g(n)(1-\delta)k}{n-1}, \frac{g(n)(1+\delta)k}{n-1}\right]\right) + o(1) \quad \text{and}$$
$$\mathbb{P}\big(\big|\deg^-(v) - k\big| \le \delta k\big) = \mu^{\langle \cdot, \mathbb{E}[\nu]\rangle}\left(\left[\frac{g(n)(1-\delta)k}{n-1}, \frac{g(n)(1+\delta)k}{n-1}\right]\right) + o(1)\,.$$

Although these results are very similar to the results that were obtained in Chapter 2 they are significant in terms of the modeling real world complex networks. In particular, many of the models for complex networks are undirected and those that have a direction tend to be acyclic [14], in contrast to the many directed cycles that exist in networks like the World Wide Web. Further, this model allows for significantly different in-degree and out-degree distributions [59]. Thus directed random dot product graphs are a highly flexible and robust model for directed complex networks.

Over the last decade or so there have been several indications that although complex networks certainly exhibit small diameter, clustering, and skewed degree distribution there are other properties that are essential to the behavior of complex networks [34, 35, 43, 44, 49]. In Chapter 4 we explore a selection of these additional properties. For instance, the ability of the physical layer of the Internet to handle the ever increasing loads placed upon it and part of the effective performance of the PageRank™ algorithm is based on the conductance and spectral gap of the underlying complex network (specifically the physical layer of the Internet and the World Wide Web). In Section 4.1 we derive an extension of the results of Chung, Lu, and Vu [22] to the conductance and spectral gap of the normalized Laplacian for a random graph with independent edges. We reduce the calculation of the second eigenvector to the calculation of the maximum eigenvalue of a deterministic matrix. We use this reduction to show that the conductance of a sufficiently dense Stochastic Kronecker Graph [40, 41] is rapidly mixing. Additionally, we use this reduction to explore the conductance and spectral properties of the random dot product graph and characterize a class of distributions μ which result in constant spectral gap and conductance. In Section 4.1 we also show, as opposed to recent results on the geometric preferential attachment [28, 29], that any cuts that would impede the rapid mixing of the random walk are essentially non-geometric. Finally, in Section 4.2 we characterize the assortativity of random dot product graphs via a technical structural lemma.

1.3 Terminology, Notation, and Technical Preliminaries

In this section we outline some basic terminology, notation and a few technical results that will be used throughout this thesis. As a general guide for graph notation we will follow the notation of Bollobás [11] and for the probabilistic results our primary reference in the work of Alon and Spencer [4].

1.3.1 Analytical Notation

In order to condense notation throughout the thesis we define some notation for some important classes of sets. First we will notate the closure of set A, that is the smallest closed set containing A, as $\texttt{closure}(A)$ and the complement of the set as $\overline{A}$. We will define the following classes of sets.

- A ball of radius r around x is $\mathbb{B}(x;r) = \left\{y \in \mathbb{R}^d \mid \|x-y\|_2 < r\right\}$.
- The cone centered around x with parameter r is

$$\mathcal{C}(x,r) = \left\{y \in \mathbb{R}^d \mid \left\langle \frac{x}{\|x\|_2}, \frac{y}{\|y\|_2} \right\rangle > 1 - r\right\}.$$

A collection of subsets $\mathcal{A} = \{A_i\}_{i\in\mathcal{I}}$ of a set X is a σ-algebra if $X \in \mathcal{A}$ and $\mathcal{A}$ is closed under complements and countable union. The Borel σ-algebra on $\mathbb{R}^d$ is the σ-algebra of subsets of $\mathbb{R}^d$ that contains the open and closed sets. We will say that a measure μ is a Borel measure if the collection of Borel sets are μ-measurable. The Lebesgue measure, denoted $\mathcal{L}$, on $\mathbb{R}^d$ is the Borel measure where $\mathcal{L}([a_1, b_1] \times \cdots \times [a_d, b_d]) = \prod_{i=1}^{d}(b_i - a_i)$ [26, 30].

For an matrix M we will define $\|M\|$ as the maximum modulus of an eigenvalue of M and define $\|M\|_\infty$ as the infinity norm of M as a matrix, that is $\|M\|_\infty = \max_{i,j} |m_{i,j}|$. Also, since every vertex in the random dot product graphs will be associated with some vector in $\mathbb{R}^d$, we will abuse notation and say that the vertex lies within a region R if the vector associated with the vertex lies within the region R. We will also define e_i as the d-dimensional vector where the i^{th} component is one and all others are zero. Further, we define $\mathbb{1} = \sum_i e_i$.

1.3.2 Graph Notation

A *graph* is an ordered pair $G = (V, E)$ where V is a set and E is a subset of the two element subsets of V. The set V is called the set of *vertices* of G, denoted $V(G)$ and the set E is the set of *edges* of G, denoted $E(G)$. A *directed graph*, or digraph, is an ordered pair (V, E) where $E \subset V \times V - \{(v, v) \mid v \in V\}$. Again, V is referred to as the set of vertices of G and E is the set of edges or arcs of the graph. If $\{u, v\} \in E(G)$ we say u and v are adjacent and denote this $u \sim v$. If $(u, v) \in E(G)$ of a directed graph G we say there is a directed edge from u to v and denote this by $u \to v$. Alternatively, we may refer to the set of edges of a directed graph as the set of arcs and say there is an arc from u to v if $(u, v) \in E(G)$. The *degree* of a vertex v is $\deg(v) = |\{u \in V(G) \mid u \sim v\}|$. The *in-degree* (respectively out-degree) of a vertex is v, denoted $\deg^+(v)$ (respectively $\deg^-(v)$), is $|\{u \in V(G) \mid u \to v\}|$ (respectively $|\{u \in V(G) \mid v \to u\}|$). A *path* between vertices s and t is a collection of distinct vertices $s = v_0, v_1, v_2, \ldots, v_k = t$ where $\{v_i, v_{i+1}\} \in E(G)$ for $0 \leq i \leq k-1$. The length of the path between s and t is one less than the number of vertices in the path. In a directed graph G a *directed path* from s to t is a collection of distinct vertices $s = v_0, v_1, v_2, \ldots, v_k = t$ where $(v_i, v_{i+1}) \in E(G)$ and the length of such path is k. A graph is *connected* if there is a path between every pair of vertices in the graph. Similarly a directed graph is *strongly connected* if there is a directed path between every ordered pair of vertices in a graph. The *diameter* of graph G, denoted $\operatorname{diam}(G)$, is the minimum k such that there is a path of length at most

k between every pair of distinct vertices, note that if a graph is not connected we say that the diameter is infinite. For a directed graph G, the *directed diameter* of the graph is the minimum k such that there is a directed path of length at of length at most k between every ordered pair of vertices. By convention, if a directed graph G is not strongly connected then the directed diameter is infinite.

The Erdős-Rényí random graph, $\mathcal{G}(n, p)$, is the graph on $[n] = \{1, 2, \ldots, n\}$ where each edge $\{i, j\}$ is present independently with probability p. A directed Erdős-Rényí random graph, $\overrightarrow{\mathcal{G}}(n, p)$, is a directed graph on the vertex set $[n]$ where each arc $i \to j$ is present independently with probability p.

1.3.3 Probabilistic Preliminaries

In this section we outline a variety of technical probabilistic inequalities that will occur throughout this thesis. We first will define a few abuses of notation we will use throughout this thesis. For a probability measure μ we will denote the expectation of a function f of a random variable generated according to μ by $\mathbb{E}[f(\mu)]$. Similarly, we denote the covariance matrix of a random variable distributed according to μ as $\operatorname{cov}(\mu)$. Note that the $\operatorname{cov}(\mu)_{ij}$ is the covariance of the i^{th} component and j^{th} component of a random variable distributed according to μ. Finally, if μ is a probability measure on $\mathbb{R}^d$, we will denote by μ_i the marginal probability distribution of μ on the i^{th} component.

In addition, we define $\mathcal{B}(n, p)$ as the binomial random variable distribution with the probability of value $k \in [n]$ being $\binom{n}{k} p^k (1-p)^{n-k}$. We will also define $\mathcal{P}(\lambda)$ as the Poisson random variable with the probability of the value $k \in \mathbb{N}$ being $\frac{\lambda^k e^{-\lambda}}{k!}$.

The following inequality regarding random variables and convex functions will be useful in the derivation of clustering for directed random dot product graphs;

Theorem 1.1 (Jensen's Inequality)**.** *If X is a random variable and ϕ is a convex function, then $\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]$.*

Although in general Chernoff Bounds give tighter bounds, there are a few cases in this thesis where the easier to derive Markov or Chebychev inequalities suffice. In fact, when summing dependent random variables the basic Chernoff Bounds we use here are not applicable, but both Markov's and Chebychev's inqualities hold.

Markov's Inequality If X is a non-negative random variable with finite expectation, then for any $\alpha > 0$, we have $\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}[X]}{\alpha}$.

Chebychev's Inequality If X is a random variable with finite second moment, then for any $\alpha > 0$ we have $\mathbb{P}(|X - \mathbb{E}[X]| > \alpha) \leq \frac{\mathrm{Var}(X)}{\alpha^2}$.

1.3.3.1 Chernoff Bounds

Throughout this thesis we will use a variety of different Chernoff bounds in order to bound the deviations from the mean of several random variables. In the hopes of creating a cleaner exposition we will state the different versions of Chernoff Bounds we use here.

C1 From [22]: Let X_i be independent random variables satisfying $|X_i| \leq M$ and let $X = \sum_i X_i$. Then

$$\mathbb{P}(|X - \mathbb{E}[X]| > a) \leq e^{-\frac{a^2}{2(\mathrm{Var}(X) + \frac{a}{3})}}.$$

C2 From [4]: Let $p_1, \ldots, p_n \in [0,1]$ with $pn = \sum_{i=1}^n p_i$. Let $X_1, \ldots, X_n$ be mutually independent with $\mathbb{P}(X_i = 1 - p_i) = p_i$ and $\mathbb{P}(X_i = -p_i) = (1 - p_i)$. Then for $\alpha > 0$,

$$\mathbb{P}(X_1 + \cdots + X_n \geq \alpha) \leq \left[\left(\frac{(1-p)}{p}\right)\left(\frac{\alpha + np}{n - (\alpha + np)}\right)\right]^{-(\alpha+pn)} \left(\frac{(1-p)n}{n - (\alpha + np)}\right)^n.$$

C2a Now by applying Chernoff Bound (C2) to $\mathcal{B}\left(n, \frac{(1-\epsilon)t}{n}\right)$, we have for $t < n$

$$\mathbb{P}\left(\mathcal{B}\left(n, \frac{(1-\epsilon)t}{n}\right) \geq t\right) \leq (1-\epsilon)^t \left(1 + \frac{\epsilon t}{n - t}\right)^{n-t} \leq ((1-\epsilon)e^{\epsilon})^t.$$

C2b Now consider

$$\begin{aligned}\mathbb{P}\left(\mathcal{B}\left(n, \frac{(1+\epsilon')\tau}{n}\right) \leq \tau\right) &= \mathbb{P}\left(n - \mathcal{B}\left(n, \frac{(1+\epsilon')\tau}{n}\right) \geq n - \tau\right) \\ &= \mathbb{P}\left(\mathcal{B}\left(n, 1 - \frac{(1+\epsilon')\tau}{n}\right) \geq n - \tau\right) \\ &= \mathbb{P}\left(\mathcal{B}\left(n, \frac{(1 - \frac{\epsilon'\tau}{n-\tau})(n-\tau)}{n}\right) \geq n - \tau\right).\end{aligned}$$

Now letting $t = n - \tau$ and $\epsilon = \frac{\epsilon'\tau}{n-\tau}$ and applying Chernoff Bound (C2a), we have

$$\mathbb{P}\left(\mathcal{B}\left(n, \frac{(1+\epsilon')\tau}{n}\right) \leq \tau\right) = \left(1 - \frac{\epsilon'\tau}{n-\tau}\right)^{n-\tau} \left(1 + \frac{\frac{\epsilon'\tau}{n-\tau}(n-\tau)}{n-(n-\tau)}\right)^{n-(n-\tau)}$$
$$= \left(1 - \frac{\epsilon'\tau}{n-\tau}\right)^{n-\tau} \left(1+\epsilon'\right)^{\tau}.$$

Now if $\tau < n$, this is at most $\left(e^{-\epsilon'}(1-\epsilon')\right)^{\tau}$.

C3 From [4]: The standard Chernoff bound: Let X_i be independent Poisson trials and let $X = \sum_i X_i$. Then for $0 < \delta < 1$,

$$\mathbb{P}(X > (1+\delta)\mathbb{E}[X]) \leq e^{-\frac{\mathbb{E}[X]\delta^2}{3}}$$
$$\mathbb{P}(X < (1-\delta)\mathbb{E}[X]) \leq e^{-\frac{\mathbb{E}[X]\delta^2}{2}}$$
$$\mathbb{P}(|X - \mathbb{E}[X]| > \delta\mathbb{E}[X]) \leq 2e^{-\frac{\mathbb{E}[X]\delta^2}{3}}.$$

C4 From [4]: Let X_i be independent Poisson trials and let $X = \sum_i X_i$ then for $0 < \delta$

$$\mathbb{P}(X > (1+\delta)\mathbb{E}[X]) \leq \left(\frac{1}{e}\left(\frac{e}{1+\delta}\right)^{(1+\delta)}\right)^{\mathbb{E}[X]} \quad [4].$$

C5 Let $X_1, X_2, \ldots, X_n$ be independent identically distributed Poisson trials and let $X = \sum_i X_i$, then by Chernoff bound (C3) for fixed $0 < \delta < 1$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \delta\mathbb{E}[X]) \leq e^{-\Theta(n)}.$$

CHAPTER II

GENERAL RANDOM DOT PRODUCT GRAPHS: SMALL WORLD PROPERTIES

In this chapter we consider the degree to which the random dot product graph model has the characteristic properties of complex networks. Specifically, we will provide a reduction for the diameter of a large class of instantiations of the random dot product graph model to the diameter of an Erdős-Rényí graph with similar average degree. We will also show that, as a result of the fundamentally geometric nature of the inner product, the random dot product graph model will exhibit positive clustering in the "friend of my friend" sense. Then, rather than explicitly proving that the random dot product graph model produces a power-law degree distribution, we will provide a "closed form" formula for the "logarithmic-histogram". That is, we will provide a formula for the probability that vertex has degree between $(1-\delta)k$ and $(1+\delta)k$, and so characterize the histogram of the degree distribution where the width of the intervals is multiplicative instead additive. In the final section of this chapter we compare the results obtained regarding random dot product graphs with the recent work of Bollobás, Jansen, and Riordan [13].

Consider the following construction which we will use to generate the random dot product graphs.

- To each vertex v in $[n]$ associate a vector X_v distributed according to μ.
- Each edge $\{i,j\}$ is present independently with probability $\frac{\langle X_i, X_j\rangle}{g(n)}$.

More formally, define the random dot product graph as follows. Let μ be a probability measure on $\mathbb{R}^d$ and $g(n) \geq 1$. The random dot product graph, denoted $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$, is the product measure $\mu^n \times (\mathcal{U}[0,1])^{\binom{n}{2}}$ together with the map that takes elements of $\left(\mathbb{R}^d\right)^n \times [0,1]^{\binom{n}{2}}$, which we shall represent as

$$(X_1, X_2, \ldots, X_n, u_{1,2}, u_{1,3}, \ldots, u_{1,n}, u_{2,3}, \ldots, u_{n-2,n}, u_{n-1,n})$$

to graphs on $[n]$ via the mapping that the edge i, j is present if and only if $\frac{\langle X_i, X_j \rangle}{g(n)} \leq u_{i,j}$. We will assume in addition that μ satisfies what we will call the *inner product condition*, which is that $(\mu \times \mu)\big(\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d \mid \langle x, y \rangle \notin [0,1]\}\big) = 0$. We note that if μ is such that $(\mu \times \mu)\big(\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d \mid \langle x, y \rangle \notin [0,c]\}\big) = 0$, then there is a measure μ', satisfying the inner product condition, defined by $\mu'(S) = \mu(\sqrt{c}S)$ such that $\mathcal{G}_{cg(n)}^{\langle \cdot,\cdot \rangle}(\mu', n)$ has the same distribution as on labeled n vertex graphs as $\mathcal{G}_g^{\langle \cdot,\cdot \rangle}(\mu, n)$. Thus the restriction of the inner product to $[0, 1]$ is no more restrictive than restricting the inner product to any interval $[0, c]$. Note that if μ satisfies the inner product condition the initial construction is the natural means to understand $\mathcal{G}_g^{\langle \cdot,\cdot \rangle}(\mu, n)$.

Now if in addition, μ satisfies that $(\mu \times \mu)\big(\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d \mid \langle x, y \rangle \notin (0,1]\}\big) = 0$ we will say that μ satisfies the *strong inner product condition*. We note that any distribution on the unit d-dimensional ball intersected with the non-negative orthant will satisfy the inner product condition, and further, if every component is strictly positive, such a distribution satisfies the strong inner product condition. With these definitions in hand, we now turn to the consideration of the diameter of $\mathcal{G}_g^{\langle \cdot,\cdot \rangle}(\mu, n)$.

2.1 Diameter

In contrast to the majority of the literature on the diameter of random graphs, we do not explicitly calculate the diameter of $\mathcal{G}_g^{\langle \cdot,\cdot \rangle}(\mu, n)$. Rather, we provide a reduction for the diameter of $\mathcal{G}_g^{\langle \cdot,\cdot \rangle}(\mu, n)$ to the diameter of a class of Erdős-Rényí graphs. To our knowledge this is the first such argument and we believe that given the current interest in the coupling of geometry and random graphs this result will provide a framework to rapidly determine the diameter of a large class of random graphs.

Theorem 2.1. *Let $r(n)$ is a decreasing function of n such that $n\mu(\mathbb{B}(0; r(n))) \to 0$ and let $g(n)$ be monotonically increasing. Suppose that $d(n)$ is a function such that for every $c, \epsilon \in (0, 1)$ there exists a c' such that $\mathbb{P}\Big(\operatorname{diam}\Big(\mathcal{G}\Big(cn, \frac{(1-\epsilon)r(n)^2}{g(n)}\Big)\Big) \leq c' d(n)\Big) = 1 - o(1)$, and μ satisfies the strong inner product condition, then $\mathbb{P}\Big(\operatorname{diam}\Big(\mathcal{G}_g^{\langle \cdot,\cdot \rangle}(\mu, n)\Big) \in \mathcal{O}(d(n))\Big) = 1 - o(1)$.*

Proof. We first note that since $n\mu(\mathbb{B}(0; r(n))) \to 0$, then with probability at least $1 - o(1)$

there are no vertices in $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ with vectors having norm less than $r(n)$. We will now proceed to show that locally $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ looks sufficiently like $\mathcal{G}\left(cn, \frac{r(n)^2}{g(n)}\right)$ for some $c \in (0, 1)$ so that the diameter results can be lifted from the $\mathcal{G}(\cdot, \cdot)$ result.

Fix $0 < \epsilon < 1$ and for each x with $\|x\| = 1$ define C_x as the cone $\mathcal{C}(x, \epsilon)$. The collection of $\{C_x\}$ forms an open cover of the d-dimensional unit sphere, which is compact by the Heine-Borel Theorem, thus there is some finite cover of d-dimensional unit sphere and hence $\mathbb{R}^d$. Let $\{C_{x_i}\}$ be the collections of sets in the finite cover such that $\mu(C_{x_i}) \neq 0$. First note that by Chernoff Bound (C5) there are at least $\frac{\mu(C_{x_i})n}{2} = c_i n$ vertices in C_{x_i} with probability at least $1-e^{-\Theta(n)}$. Furthermore for any two vertices in C_{x_i}, the probability of an edge between them is at least $\frac{r(n)^2(1-\epsilon)}{g(n)}$. Letting n_i be the number of vertices in C_{x_i}, we couple $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ restricted to C_{x_i} with the graph $\mathcal{G}\left(n_i, \frac{r(n)^2(1-\epsilon)}{g(n)}\right)$. Since every edge in $\mathcal{G}\left(n_i, \frac{r(n)^2(1-\epsilon)}{g(n)}\right)$ occurs with probability that is less than the corresponding edge in $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$, then $\mathcal{G}\left(n_i, \frac{r(n)^2(1-\epsilon)}{g(n)}\right)$ is, via the coupling, a subgraph of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$. In particular, $\text{diam}\left(\mathcal{G}\left(n_i, \frac{r(n)^2(1-\epsilon)}{g(n)}\right)\right)$ is greater than the diameter of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ restricted to C_{x_i}. Thus we restrict our attention to the diameter of $\mathcal{G}\left(n_i, \frac{r(n)^2(1-\epsilon)}{g(n)}\right)$. Since n_i is at least $c_i n$ with high probability, and since $g(n)$ and $r(n)$ are monotonic, the asymptotic behavior of the diameter of $\mathcal{G}\left(n_i, \frac{r(n)^2(1-\epsilon)}{g(n)}\right)$ is no worse than the asymptotic behavior of $\mathcal{G}\left(n_i, \frac{r(\frac{n_i}{c_i})^2(1-\epsilon)}{g(\frac{n_i}{c_i})}\right)$. Now, by assumption there is some $c_i' = c_i'(\epsilon)$ so that $\mathbb{P}\left(\mathcal{G}\left(n_i, \frac{r\left(\frac{n_i}{c_i}\right)^2(1-\epsilon)}{g\left(\frac{n_i}{c_i}\right)}\right) \leq c_i' d(n_i)\right) = 1 - o(1)$. Thus with high probability the diameter $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ restricted to C_{x_i} is at most $c_i' d(n)$.

We now consider the edges present between cones C_{x_i} and C_{x_j}. Since μ satisfies the strong inner product condition, there are regions $\mathcal{R}_i \subseteq C_{x_i}$ and $\mathcal{R}_j \subseteq C_{x_j}$ with positive measure, together with a constant $c_{ij}(\epsilon) = c_{ij} > 0$ such that for any $x_i \in \mathcal{R}_\rangle$ and $x_j \in \mathcal{R}_j$, $\langle x_i, x_j\rangle \geq c_{ij}$. Thus, the expected number of edges between C_{x_i} and C_{x_j} is at least $\frac{c_{ij}(c_i c_j)n^2}{g(n)}$. Now if $g(n)$ is $o\left(n^2\right)$ this expectation goes to infinity, hence there is almost surely an edge between $\mathcal{R}_i$ and $\mathcal{R}_j$, and thus between C_{x_i} and C_{x_j}.

Since there are only finitely many cones C_{x_i} under consideration, there exists $c' = \max c_i'$ such that the diameter of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ restricted to C_{x_i} is at most $c'd(n)$ almost surely. Then, combining this with the fact that there is almost surely an edge between every two cones, yields that the diameter of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu,n)$ is almost surely at most $2c'd(n)+1$. □

We observe that the condition that $n\mu(\mathbb{B}(0;r(n))) \to 0$ implicitly restricts the distributions μ for which this reduction works. If $r(n)$ approaches 0 too fast the Erdős-Rényí resulting from this reduction is not connected. In other words, if μ has enough mass clustered around the origin, this reduction scheme will not yield a proof of connectivity let alone diameter. However, we feel that this is reasonable, as if too many vertices have small norm then we would expect the random dot product graph to be disconnected. Also, we note that, as will be the case for degree distribution, there is an one-dimensional distribution that plays a key role in the analysis of the diameter $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu,n)$. In particular, define the measure $\mu^{\|\cdot\|_2}(S) = \int_{\left\{X\in\mathbb{R}^d \mid \|X\|_2\in S\right\}} d\mu(X)$. Then the condition that $n\mu(\mathbb{B}(0;r(n))) \to 0$ can be expressed as $\mu^{\|\cdot\|_2}([0,r(n)))$ is $o\left(\frac{1}{n}\right)$. Thus the diameter of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu,n)$ is heavily influenced by a one-dimensional aspect of μ, rather than relying on all d-dimensions.

It is worth noting at this point that this methodology for proving the reduction will in fact prove a larger class of results. For instance, by relaxing the condition that $\mu(\mathbb{B}(0;r(n)))$ is $o\left(\frac{1}{n}\right)$ to $\frac{n\mu(\mathbb{B}(0;r(n)))}{f(n)} \to 0$, the same methodology will yield that there is a subgraph excluding at most $o(f(n))$ vertices that has diameter on the order of an appropriate Erdős-Rényí graph with high probability. In fact, by carefully constructing and analyzing the cover by cones for a specific distribution, it is possible to loosen the strong inner product condition constraint and derive tighter results. For instance, consider the distribution ρ where e_1, e_2 and $\frac{1}{2}e_1 + \frac{1}{2}e_2$ have equal weight, and all other vectors have weight 0. The distribution ρ clearly does not satisfy the strong inner product condition, but by choosing the appropriate cones, one can easily see that the diameter of $\mathcal{G}^{\langle\cdot,\cdot\rangle}g(\rho,n)$ is at most $\operatorname{diam}\left(\mathcal{G}\left(\frac{n}{4},\frac{1}{g(n)}\right)\right) + \operatorname{diam}\left(\mathcal{G}\left(\frac{n}{4},\frac{1}{\sqrt{2}g(n)}\right)\right) + 2$.

Furthermore, theses techniques can also be used to prove tighter results based on the relationship between $r(n)$ and the diameter. For instance, if $\frac{nr(n)}{g(n)}$ is $\omega(\log(n))$ then instead of reducing to diameter of Erdős-Rényí graphs of the form $\mathcal{G}\left(cn, c'\frac{r(n)^2}{g(n)}\right)$, it can be instead shown that the diameter reduces to the behavior of $\mathcal{G}\left(cn, c'\frac{r(n)}{g(n)}\right)$. We believe that a careful analysis which conditions on the behavior of $\frac{nr(n)}{g(n)}$ may lead to a series of finer results reducing the diameter to the behavior of $\mathcal{G}\left(cn, c' \frac{r(n)}{g(n)}\right)$.

2.2 Clustering

Clustering is the most natural of the three hallmark properties, especially in social networks, and therefore it is key from an applications point of view to characterize the clustering of $\mathcal{G}_g^{\langle \cdot,\cdot \rangle}(\mu, n)$. In fact, in contrast to the surprise in the analytical community when power-law degree distribution and small world properties were discovered, when the analysis revealed that social networks exhibited clustering there was little surprise. Perhaps this is because, although there is no a priori reason to expect for a large network to have a highly skewed degree distribution or a small diameter, from our personal experience we can readily explain away clustering. That is, we know that we are more likely to be friends with our friend's friends than a random other person, which results in clustering in the social network. For this reason, although alternative definitions of clustering have been proposed, we focus on the "friend-of-my-friend" clustering. That is, we show that for any set of vertices u, v, w and all nontrivial distributions μ, we have $\mathbb{P}(u \sim w \mid u \sim v, v \sim w) > \mathbb{P}(u \sim w)$ in $\mathcal{G}_g^{\langle \cdot,\cdot \rangle}(\mu, n)$.

Theorem 2.2. *Let u, v, w be vertices in $\mathcal{G}_g^{\langle \cdot,\cdot \rangle}(\mu, n)$. Then $\mathbb{P}(u \sim w \mid u \sim v, v \sim w) \geq \mathbb{P}(u \sim w)$. Furthermore, this inequality is tight if and only if there exists some point $x \in \mathbb{R}^d$ so that $\mu(\{x\}) = 1$.*

Proof. First note that $\mathbb{P}(u \sim v) = \mathbb{P}(u \sim w) = \mathbb{P}(v \sim w)$ and furthermore since X_u and X_v are independent, $g(n)\mathbb{P}(u \sim v) = \langle \mathbb{E}[\mu], \mathbb{E}[\mu] \rangle = \|\mathbb{E}[\mu]\|_2^2$. Similarly

$$g(n)^2 \mathbb{P}(u \sim v, v \sim w) = \mathbb{E}[\mu]^T \, \mathbb{E}\left[\mu\mu^T\right] \mathbb{E}[\mu].$$

Now in order to show that $\mathbb{P}(u \sim w \mid u \sim v, v \sim w) \geq \mathbb{P}(u \sim w)$ it suffices to show that

$$\mathbb{P}(u \sim w, w \sim v, v \sim w) - \mathbb{P}(u \sim w)\, \mathbb{P}(u \sim v, v \sim w) \geq 0.$$

So now consider

$$\begin{aligned} &g(n)^3 \left(\mathbb{P}(u \sim v, v \sim w, w \sim u) - \mathbb{P}(u \sim w)\, \mathbb{P}(u \sim v, v \sim w)\right) \\ &= \int X^T \mathbb{E}\left[\mu\mu^T\right]^2 X d\mu(X) - \|\mathbb{E}[\mu]\|_2^2 \, \mathbb{E}[\mu]^T \, \mathbb{E}\left[\mu\mu^T\right] \mathbb{E}[\mu]. \end{aligned}$$

By considering the integrand component by component (indexed by i, j, and k), we have that this is

$$
\begin{aligned}
&= \sum_{i,j,k} \left(\int X_i \mathbb{E}\left[\mu_i \mu_j\right] \mathbb{E}\left[\mu_j \mu_k\right] X_k d\mu(X) - \mathbb{E}\left[\mu_i\right]^2 \mathbb{E}\left[\mu_j\right] \mathbb{E}\left[\mu_j \mu_k\right] \mathbb{E}\left[\mu_k\right] \right) \\
&= \sum_{i,j,k} \Big(\mathbb{E}\left[\mu_i \mu_k\right] \mathbb{E}\left[\mu_i \mu_j\right] \mathbb{E}\left[\mu_j \mu_k\right] - \mathbb{E}\left[\mu_i\right]^2 \mathbb{E}\left[\mu_j\right] \mathbb{E}\left[\mu_j \mu_k\right] \mathbb{E}\left[\mu_k\right] \Big) \\
&= \sum_{i,j,k} \mathbb{E}\left[\mu_j \mu_k\right] \Big(\mathbb{E}\left[\mu_i \mu_j\right] \mathbb{E}\left[\mu_i \mu_k\right] - \mathbb{E}\left[\mu_i\right]^2 \mathbb{E}\left[\mu_j\right] \mathbb{E}\left[\mu_k\right] \Big).
\end{aligned}
$$

Now let X be distributed as μ and let Y be an arbitrary vector in $\mathbb{R}^d$. Then $Y^T \operatorname{cov}(X)\, Y = Y^T \mathbb{E}\left[X X^T\right] Y - Y^T \mathbb{E}\left[X\right]^T \mathbb{E}\left[X\right] Y = \mathbb{E}\left[\langle Y, X\rangle^2\right] - \mathbb{E}\left[\langle Y, X\rangle\right]^2 = \operatorname{Var}(\langle Y, X\rangle) \geq 0$. Thus $\operatorname{cov}(\mu)$ is a symmetric positive definite matrix. Thus there is some orthogonal matrix Q such that $Q \operatorname{cov}(\mu)\, Q^T$ is diagonal. We also observe that $\langle Q X_i, Q X_j\rangle = (Q X_i)^T (Q X_j) = X_i^T Q^T Q X_j = \langle X_i, X_j\rangle$ and further

$$
\operatorname{cov}(Q\mu) = \mathbb{E}\left[Q \mu \mu^T Q\right] - \mathbb{E}\left[Q\mu\right] \mathbb{E}\left[Q\mu\right]^T = Q \operatorname{cov}(\mu)\, Q^T.
$$

Thus we may assume without loss of generality that $\operatorname{cov}(\mu)$ is diagonal, and in particular if $i \neq j$, then $\mathbb{E}\left[\mu_i \mu_j\right] = \mathbb{E}\left[\mu_i\right] \mathbb{E}\left[\mu_j\right]$. Hence if $j, k \neq i$,

$$
\mathbb{E}\left[\mu_i \mu_j\right] \mathbb{E}\left[\mu_i \mu_k\right] - \mathbb{E}\left[\mu_i\right]^2 \mathbb{E}\left[\mu_j\right] \mathbb{E}\left[\mu_k\right] = 0.
$$

Furthermore, if $i = k$ and $i \neq j$, then

$$
\mathbb{E}\left[\mu_i \mu_j\right] \mathbb{E}\left[\mu_i \mu_k\right] - \mathbb{E}\left[\mu_i\right]^2 \mathbb{E}\left[\mu_j\right] \mathbb{E}\left[\mu_k\right] = \mathbb{E}\left[\mu_j\right] \mathbb{E}\left[\mu_i\right] \operatorname{Var}(\mu_i),
$$

and similarly for $i = j$ and $i \neq k$. Finally if $i = j = k$, then

$$
\mathbb{E}\left[\mu_i \mu_j\right] \mathbb{E}\left[\mu_i \mu_k\right] - \mathbb{E}\left[\mu_i\right]^2 \mathbb{E}\left[\mu_j\right] \mathbb{E}\left[\mu_k\right] = \mathbb{E}\left[\mu_i^2\right]^2 - \mathbb{E}\left[\mu_i\right]^4 = \operatorname{Var}(\mu_i) \left(\mathbb{E}\left[\mu_i^2\right] + \mathbb{E}\left[\mu_i\right]^2 \right).
$$

Thus

$$
\begin{aligned}
&g(n)^3 \left(\mathbb{P}(u \sim v, v \sim w, w \sim u) - \mathbb{P}(u \sim w)\, \mathbb{P}(u \sim v, v \sim w)\right) \\
&= \sum_i \mathbb{E}\left[\mu_i^2\right] \operatorname{Var}(\mu_i) \left(\mathbb{E}\left[\mu_i^2\right] + \mathbb{E}\left[\mu_i\right]^2 \right) + 2 \sum_{i \neq j} \mathbb{E}\left[\mu_i\right]^2 \operatorname{Var}(\mu_i)\, \mathbb{E}\left[\mu_j\right]^2 \\
&\geq 0.
\end{aligned}
$$

Now if there exists some $x \in \mathbb{R}^d$ so that $\mu(\{x\}) = 1$, then $\text{Var}(\mu_i) = 0$ for all i and $\mathbb{P}(u \sim w \mid u \sim v, v \sim w) = \mathbb{P}(u \sim w)$. Conversely, if $\mathbb{P}(u \sim w \mid u \sim v, v \sim w) = \mathbb{P}(u \sim w)$, then $\text{Var}(\mu_i) = 0$ for all i. But then, $\mu_i(\{\mathbb{E}[\mu_i]\}) = 1$ and in particular $\mu(\{\mathbb{E}[\mu]\}) = 1$. □

In other words, the $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ model exhibits positive clustering in the "friend-of-my-friend" sense unless $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ is an Erdős-Rényí random graph, in which case we know that since presence of each edge is independent and there is no clustering. Looking at the nature of the proof, it is apparent that the covariance matrix of μ, will play a fundamental role in determining the clustering of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$. In fact, observing that the covariance matrix for an Erdős-Rényí graph will be identically 0, it is reasonable to presume that the covariance matrix in some manner quantifies the amount of clustering present. That is, if the some norm-like quantity for the covariance matrix is large, then there is a corresponding increase in the clustering coefficient. Although we do not explicitly have such a result currently, such a result would be extremely interesting in terms of tailoring the clustering coefficient of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ to particular real world networks.

2.3 Degree Distribution

In this section we analyze the degree distribution of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ in terms of an alternative measure that is derived from μ. In particular, if X is a random variable distributed according to μ, we concern ourselves with the distribution of $\langle X, \mathbb{E}[\mu]\rangle$. To that end, in addition to defining $\mathcal{L}$ as the Lebesgue measure, we define the following measures:

- $\mu^{\langle\cdot,\cdot\rangle}(S) = \iint_{\langle \mathbb{E}[\mu], X_v\rangle \in S} d\mu(X_v)$ and
- $\mu_g^{\langle\cdot,\cdot\rangle}(S) = \iint_{\frac{\langle \mathbb{E}[\mu], X_v\rangle}{g(n)} \in S} d\mu(X_v)$.

Before turning to the primary result of this section, we need the following lemma analyzing the behavior of a binomial random variable, $\mathcal{B}(n, p)$, where p is a random variable.

Lemma 2.1. *Let P be a random variable taking values in $[0, 1]$. Let $0 < \delta, \epsilon < 1$ be such that $(1 + \delta)(1 - \epsilon) > 1$. Let $f(\epsilon, \delta, k) = ((1 + \epsilon)e^{-\epsilon})^{(1+\delta)k} + ((1 - \epsilon)e^{\epsilon})^{(1-\delta)k}$ and*

$\hat{f}(\epsilon,\delta,k) = ((1+\epsilon)e^{-\epsilon})^{(1-\delta)k} + ((1-\epsilon)e^{\epsilon})^{(1+\delta)k}$. *Then, for* $0 \le k \le n$

$$(1-f(\epsilon,\delta,k))\,\mathbb{P}\!\left(\frac{(1+\epsilon)(1-\delta)k}{n} \le P \le \frac{(1-\epsilon)(1+\delta)k}{n}\right) \le \mathbb{P}(|\mathcal{B}(n,P)-k| \le \delta k) \quad \textit{and}$$
$$\hat{f}(\epsilon,\delta,k) + \mathbb{P}\!\left(\frac{(1-\epsilon)(1-\delta)k}{n} \le P \le \frac{(1+\epsilon)(1+\delta)k}{n}\right) \ge \mathbb{P}(|\mathcal{B}(n,P)-k| \le \delta k)\,.$$

Proof. Define the following events

$$\begin{aligned}
\mathcal{E}_1 &= \left\{P < \frac{(1-\epsilon)(1-\delta)k}{n}\right\}\\
\mathcal{E}_2 &= \left\{\frac{(1-\epsilon)(1-\delta)k}{n} \le P < \frac{(1+\epsilon)(1-\delta)k}{n}\right\}\\
\mathcal{E}_3 &= \left\{\frac{(1+\epsilon)(1-\delta)k}{n} \le P \le \frac{(1-\epsilon)(1+\delta)k}{n}\right\}\\
\mathcal{E}_4 &= \left\{\frac{(1-\epsilon)(1+\delta)k}{n} < P \le \frac{(1+\epsilon)(1+\delta)k}{n}\right\}\\
\mathcal{E}_5 &= \left\{\frac{(1+\epsilon)(1+\delta)k}{n} < P\right\}.
\end{aligned}$$

Note that by the choice of ϵ and δ these events are disjoint.

Now we observe that

$$\begin{aligned}
\mathbb{P}(|\mathcal{B}(n,P)-k| \le \delta k) &= \mathbb{P}(|\mathcal{B}(n,P)-k| \le \delta k \mid \mathcal{E}_1) + \mathbb{P}(|\mathcal{B}(n,P)-k| \le \delta k \mid \mathcal{E}_2 \cup \mathcal{E}_3 \cup \mathcal{E}_4)\\
&\quad + \mathbb{P}(|\mathcal{B}(n,P)-k| \le \delta k \mid \mathcal{E}_5)\\
&\le \mathbb{P}(\mathcal{B}(n,P) \ge (1-\delta)k \mid \mathcal{E}_1)\\
&\quad + \mathbb{P}(\mathcal{E}_2 \cup \mathcal{E}_3 \cup \mathcal{E}_4) + \mathbb{P}(\mathcal{B}(n,P) \le (1+\delta)k \mid \mathcal{E}_5)\\
&\le \mathbb{P}\!\left(\mathcal{B}\!\left(n, \frac{(1-\epsilon)(1-\delta)k}{n}\right) \ge (1-\delta)k\right)\\
&\quad + \mathbb{P}\!\left(\frac{(1-\epsilon)(1-\delta)k}{n} \le P \le \frac{(1+\epsilon)(1+\delta)k}{n}\right)\\
&\quad + \mathbb{P}\!\left(\mathcal{B}\!\left(n, \frac{(1+\epsilon)(1+\delta)k}{n}\right) \le (1+\delta)k\right)\\
&\le f(\epsilon,\delta,k) + \mathbb{P}\!\left(\frac{(1-\epsilon)(1-\delta)k}{n} \le P \le \frac{(1+\epsilon)(1+\delta)k}{n}\right).
\end{aligned}$$

Where the last inequality comes from applying the Chernoff Bounds (C2a) and (C2b), respectively.

For the other direction, we note that

$$\begin{aligned}
\mathbb{P}(|\mathcal{B}(n,P)-k| \le \delta k) &\ge \mathbb{P}(\mathcal{E}_3)\,\mathbb{P}(|\mathcal{B}(n,P)-k| \le \delta k \mid \mathcal{E}_3) \\
&\ge \mathbb{P}(\mathcal{E}_3) - \mathbb{P}(\mathcal{E}_3)\,\mathbb{P}(\mathcal{B}(n,P) \ge (1+\delta)k \mid \mathcal{E}_3) \\
&\quad - \mathbb{P}(\mathcal{E}_3)\,\mathbb{P}(\mathcal{B}(n,P) \le (1-\delta)k \mid \mathcal{E}_3) \\
&\ge \mathbb{P}(\mathcal{E}_3) - \mathbb{P}(\mathcal{E}_3)\,\mathbb{P}\bigg(\mathcal{B}\bigg(n, \frac{(1-\epsilon)(1+\delta)k}{n}\bigg) \ge (1+\delta)k\bigg) \\
&\quad - \mathbb{P}(\mathcal{E}_3)\,\mathbb{P}\bigg(\mathcal{B}\bigg(n, \frac{(1+\epsilon)(1-\delta)k}{n}\bigg) \le (1-\delta)k\bigg) \\
&\ge \Big(1 - \hat{f}(\epsilon,\delta,k)\Big)\,\mathbb{P}\bigg(\frac{(1+\epsilon)(1-\delta)k}{n} \le P \le \frac{(1-\epsilon)(1+\delta)k}{n}\bigg).
\end{aligned}$$

Where the last inequality comes from the application of the Chernoff Bounds (C2a) and (C2b), respectively. ☐

With this in hand, we will characterize the "log-histogram" of the degree distribution in terms of $\mu^{\langle\cdot,\cdot\rangle}$. Thus, by determining the behavior of the one-dimensional distribution $\mu^{\langle\cdot,\cdot\rangle}$, we can specify the degree distribution. By considering rotations of $[0,1]$ into $\mathbb{R}^d$, this is clearly a means of constructing μ so that $\mu^{\langle\cdot,\cdot\rangle} = \rho$ for any ρ which is a distribution on $[0,1]$. Furthermore, this flexibility allows the model to successfully deal with the possibility that the distribution of degrees for many complex networks are not truly power-laws, but rather are simply "near" a power-law distribution. In particular, consider the case of the degree distribution for the autonomous system layer of the Internet. Despite the results of Faloutsos, Faloutsos, and Faloutsos indicating the degree distribution for this network [27] there has been recent work indicating that the portion of their methodology using trace-route sampling may be inherently biased towards revealing a power-law. In fact, Lakhina, Byers, Crovella, and Xie show experimentally that traceroute sampling on an Erdős-Rényí random graph yields power-law like behavior [39]. Achlioptas, Clauset, Kempe, and Moore confirm this result theoretically by showing that d-regular random graphs can exhibit power-law degree sequence under traceroute sampling [2].

Theorem 2.3. *Suppose $\mu^{\langle\cdot,\cdot\rangle}$ is a Borel measure such that for any Borel-measurable set S there exists a c such that $\mu^{\langle\cdot,\cdot\rangle}(S) \le c\mathcal{L}(S)$. If $g(n)$ is $o(n)$, then for any vertex v in*

$\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ *and any fixed* $0 < \delta < 1$,

$$\mathbb{P}(|\deg(v) - k| \leq \delta k) = \mu^{\langle\cdot,\cdot\rangle}\left(\left[\frac{g(n)(1-\delta)k}{n-1}, \frac{g(n)(1+\delta)k}{n-1}\right]\right) + o(1)\,.$$

Further, if k *is* $\omega(1)$*, then error term can be tightened to* $\mathcal{O}\left(\sqrt{\frac{\ln(k)}{k}}\right)$.

Proof. We first note that

$$\begin{aligned}
\mathbb{P}(\deg(v) = k) &= \iint \sum_{\substack{S\subset V\\|S|=k}} \prod_{s\in S} \frac{\langle X_s, X_v\rangle}{g(n)} \prod_{t\notin S}(1 - \frac{\langle X_t, X_v\rangle}{g(n)})d\mu(X_v)d\mu(X_s)\cdots d\mu(X_t)\\
&= \int \sum_{\substack{S\subset V\\|S|=k}} \prod_{s\in S} \frac{\langle \mathbb{E}\left[\mu\right], X_v\rangle}{g(n)} \prod_{t\notin S}(1 - \frac{\langle \mathbb{E}\left[\mu\right], X_v\rangle}{g(n)})d\mu(X_v)\\
&= \int \binom{n-1}{k}\left(\frac{\langle \mathbb{E}\left[\mu\right], X_v\rangle}{g(n)}\right)^k\left(1 - \frac{\langle \mathbb{E}\left[\mu\right], X_v\rangle}{g(n)}\right)^{n-1-k} d\mu(X_v)\\
&= \int \mathbb{P}\left(\mathcal{B}\left(n-1, \frac{\langle \mathbb{E}\left[\mu\right], X_v\rangle}{g(n)}\right) = k\right) d\mu(X_v)
\end{aligned}$$

Thus if P is a random variable distributed according to $\mu_g^{\langle\cdot,\cdot\rangle}$, then

$$\mathbb{P}(|\deg(v) - k| \leq \delta k) = \mathbb{P}(|\mathcal{B}(n-1, P) - k| \leq \delta k)\,.$$

Thus by Lemma 2.1, for $0 < \delta, \epsilon < 1$ and $(1+\delta)(1-\epsilon) > 1$.

$$\begin{aligned}
&\mathbb{P}(|\deg(v) - k| \leq \delta k) \leq f(\epsilon, k, \delta) + \mu_g^{\langle\cdot,\cdot\rangle}\left(\left[\frac{(1-\delta)(1-\epsilon)k}{n-1}, \frac{(1+\delta)(1+\epsilon)k}{n-1}\right]\right) \quad \text{and}\\
&\mathbb{P}(|\deg(v) - k| \leq \delta k) \geq \left(1 - \hat{f}(\epsilon, k, \delta)\right)\mu_g^{\langle\cdot,\cdot\rangle}\left(\left[\frac{(1-\delta)(1+\epsilon)k}{n-1}, \frac{(1+\delta)(1-\epsilon)k}{n-1}\right]\right)
\end{aligned}$$

Consider then the case where k is $\omega(1)$. Now if $\epsilon = \sqrt{\frac{\ln(k)}{(1-\delta)k}}$, then

$$\begin{aligned}
\ln\left(\sqrt{\frac{k}{\ln(k)}}\left((1+\epsilon)e^{-\epsilon}\right)^{(1-\delta)k}\right) &= \frac{1}{2}\ln(k) - \frac{1}{2}\ln(\ln(k)) + (1-\delta)k\left(\epsilon + \ln(1-\epsilon)\right)\\
&\leq \frac{1}{2}\ln(k) - \frac{1}{2}\ln(\ln(k)) - (1-\delta)k\frac{\epsilon^2}{2}\\
&= -\frac{1}{2}\ln(\ln(k)) \to -\infty.
\end{aligned}$$

Further, we have that

$$\ln\left(\sqrt{\frac{k}{\ln(k)}}\left((1-\epsilon)e^{\epsilon}\right)^{(1-\delta)k}\right) = \frac{1}{2}\ln(k) - \frac{1}{2}\ln(\ln(k)) + (1-\delta)k\left(-\epsilon + \ln(1+\epsilon)\right)$$
$$\leq \frac{1}{2}\ln(k) - \frac{1}{2}\ln(\ln(k)) - (1-\delta)k\left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}\right)$$
$$= -\frac{1}{2}\ln(\ln(k)) + \sqrt{\frac{\ln^3(k)}{(1-\delta)k}} \to -\infty.$$

Thus $f(\epsilon,\delta,k)$ and $\hat{f}(\epsilon,\delta,k)$ are $o\left(\sqrt{\frac{k}{\ln(k)}}\right)$ and hence

$$\left(1 - o\left(\sqrt{\frac{\ln(k)}{k}}\right)\right)\mu_g^{\langle\cdot,\cdot\rangle}\left(\left[\frac{(1-\delta)(1+\epsilon)k}{n-1}, \frac{(1+\delta)(1-\epsilon)k}{n-1}\right]\right) \leq \mathbb{P}(|\deg(v) - k| \leq \delta k)$$
$$o\left(\sqrt{\frac{\ln(k)}{k}}\right) + \mu_g^{\langle\cdot,\cdot\rangle}\left(\left[\frac{(1-\delta)(1-\epsilon)k}{n-1}, \frac{(1+\delta)(1+\epsilon)k}{n-1}\right]\right) \geq \mathbb{P}(|\deg(v) - k| \leq \delta k)\,.$$

At this point we observe that $\mu_g^{\langle\cdot,\cdot\rangle}(S) = \mu^{\langle\cdot,\cdot\rangle}\left(\frac{1}{g(n)}S\right)$, and so

$$\left(1 - o\left(\sqrt{\frac{\ln(k)}{k}}\right)\right)\mu^{\langle\cdot,\cdot\rangle}\left(\left[(1-\delta)(1+\epsilon)t_n^k, (1+\delta)(1-\epsilon)t_n^k\right]\right) \leq \mathbb{P}(|\deg(v) - k| \leq \delta k)$$
$$o\left(\sqrt{\frac{\ln(k)}{k}}\right) + \mu^{\langle\cdot,\cdot\rangle}\left(\left[(1-\delta)(1-\epsilon)t_n^k, (1+\delta)(1+\epsilon)t_n^k\right]\right) \geq \mathbb{P}(|\deg(v) - k| \leq \delta k)$$

where $t_n^k = \frac{kg(n)}{n-1}$. We now observe that

$$\mu^{\langle\cdot,\cdot\rangle}\left(\left[(1-\delta)(1-\epsilon)t_n^k, (1+\delta)(1+\epsilon)t_n^k\right]\right) - \mu^{\langle\cdot,\cdot\rangle}\left(\left[(1-\delta)(1+\epsilon)t_n^k, (1+\delta)(1-\epsilon)t_n^k\right]\right)$$

is at most $c\frac{4\epsilon k g(n)}{n-1}$ Which by the choice of ϵ is $\mathcal{O}\left(\frac{\sqrt{k\ln(k)}g(n)}{n}\right)$. But if k is $\mathcal{O}\left(\frac{n}{g(n)}\right)$, then this is $\mathcal{O}\left(\sqrt{\frac{\ln(k)}{k}}\right)$.

Now consider the case when k is $\omega\left(\frac{n}{g(n)}\right)$. Then, since the support of $\mu^{\langle\cdot,\cdot\rangle}$ is bounded, there is some c' so that

$$\mathbb{P}((1-\delta)k \leq \deg(v)) \leq \mathbb{P}\left(\mathcal{B}\left(n-1, \frac{c'}{g(n)}\right) \geq (1-\delta)k\right)$$
$$\leq \left(\frac{1}{e}\left(\frac{e}{\frac{(1-\delta)kg(n)}{c'(n-1)}}\right)^{\frac{(1-\delta)kg(n)}{n-1}}\right)^{\frac{c'(n-1)}{g(n)}}$$
$$= e^{-\frac{c'(n-1)}{g(n)}}\left(\frac{ec'(n-1)}{(1-\delta)kg(n)}\right)^{(1-\delta)k}.$$

where the second inequality follows from Chernoff Bound (C4). Now since $\frac{(n-1)}{kg(n)} \to 0$,

$$\sqrt{\frac{k}{\ln(k)}} e^{-\frac{c(n-1)}{g(n)}} \left(\frac{ec(n-1)}{(1-\delta)kg(n)} \right)^{(1-\delta)k} \to 0.$$

Thus $\mathbb{P}((1-\delta)k \leq \deg(v) \leq (1+\delta)k)$ is $\mathcal{O}\left(\sqrt{\frac{\ln(k)}{k}}\right)$ if k is $\omega\left(\frac{n}{g(n)}\right)$.

Now we consider the case where k is constant. Then we have that

$$\begin{aligned}
\mathbb{P}((1-\delta)k \leq \deg(v) \leq (1+\delta)k) &\leq \mathbb{P}\left(\deg(v) \leq (1+\delta)\sqrt{\frac{n}{g(n)}}\right) \\
&\leq \mu^{\langle\cdot,\cdot\rangle}\left(\left[0, \sqrt{\frac{g(n)}{n}}\right]\right) + \mathcal{O}\left(\sqrt{\frac{\ln(\frac{n}{g(n)})}{\frac{n}{g(n)}}}\right) \\
&\leq \sqrt{\frac{g(n)}{n}} + \mathcal{O}\left(\sqrt{\frac{\ln(\frac{n}{g(n)})}{\frac{n}{g(n)}}}\right).
\end{aligned}$$

This is $o(1)$ as desired. □

At this point, we observe that if $\mu^{\langle\cdot,\cdot\rangle}$ is skewed, this will result in the degree distribution of a vertex being skewed as well. That is, the degree distribution of a vertex has a similar shape as the distribution $\mu^{\langle\cdot,\cdot\rangle}$. Thus with appropriate choice of distribution μ, we have that $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ exhibits all three primary properties of complex networks; clustering, small diameter, and skewed degree distribution.

2.4 Non-homogeneous Random Graphs and $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$

Independently, Bollobás, Jansen, and Riordan [13] recently developed a general model for sparse inhomogeneous random graphs. Their model and $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ both lie in the same framework random graphs where the edges are present independently, but with potentially different probabilities. We refer to such graphs as of non-homogeneous random graphs. At this point, it will be enlightening to clarify the differences and similarities between the two models and indicate a few areas of extension for both models. In their paper Bollobás et al. detail several different slight variations on the core random graph model under consideration. In order to make the comparison cleaner, we will focus only one such variation, although in principle similar comments should hold for all of the variations. With that in mind, we cast the model of inhomogeneous random graphs in the same framework as $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$

as follows. Let μ be a Borel probability measure on a separable metric space $\mathcal{S}$ and let a kernel function κ be a symmetric non-negative function on $\mathcal{S}\times\mathcal{S}$. In addition, Bollobás et al. assume certain technical "smoothness" constraints for κ. Then the inhomogeneous random graph on n vertices is the product measure μ^n coupled with the map that takes elements of $\mathcal{S}^n \times (\mathcal{U}[0,1])^{\binom{n}{2}}$, represented as $(X_1, X_2, \ldots, X_n, u_{1,2}, u_{1,3}, \ldots, u_{1,n}, u_{2,3}, \ldots, u_{n-2,n}, u_{n-1,n})$ to graphs on $[n]$ via the mapping where the edge $\{i,j\}$ is present if and only if $u_{i,j} \leq \frac{\kappa(X_i,X_j)}{n}$. We note that this expression makes it clear that $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu,n)$ and inhomogeneous random graphs are two aspects of the same general random graph model. Whereas $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu,n)$ limits the scope of random graphs under considerations by fixing the kernel function κ as $\langle\cdot,\cdot\rangle$, the inhomogeneous random graphs are limited by the fixing of $g(n)$ as n. Additionally, Bollobás, et al. assume concentration of the degree sequence. We note that many of the arguments for $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu,n)$ hold only for $g(n)$ being $o(1)$ and the results Bollobás et al. rely heavily on $g(n) = n$ and the assumption on the concentration of degrees. Thus, although they are comparable random graph models there is in essence no means of setting the parameters of inhomogeneous random graphs and $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu,n)$ to yield the same random graph model. Rather, there respective generalizations yield the same class of random graph models. Despite this lack of overlap, it is insightful to compare the results for diameter, clustering, and distribution.

In their analysis of the diameter of the inhomogeneous random graphs, Bollobás et al. turn to the method of branching process that had previously been extremely successful in analyzing the diameter behavior of the Erdős-Rényí random graphs [9, 10, 12], especially near the phase-transition, which is the natural regime for the inhomogeneous random graph. Contrast this to our methodology, which instead of explicitly calculating the diameter for the large range of parameters under consideration, provides an explicit reduction to the diameter of a related class of Erdős-Rényí random graphs. Thus, in a meta-mathematical sense, our methodology uses branching processes in that a large portion of the results for the diameter of Erdős-Rényí random graphs relying on branching processes. We feel that for the case of arbitrary average degree and general kernel functions a combination of these methods will be most useful. For kernel functions that are sufficiently like the inner product over a compact metric space, we feel that the compactness reduction argument we use will provide the cleanest argument. However, for even more general kernel functions, it is not immediate how this compactness argument may be extended and thus it may be necessary to revert to the branching process methodology.

Although clustering is an important property in the study of random networks as models of complex networks, it doesn't carry the same importance in the study of general random graphs. Thus it is unsurprising that the paper of Bollobás et al. fails to address the clustering of the inhomogeneous random graphs. Fortuitously, the argument of regarding the clustering of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ places no restriction on $g(n)$ and thus we know there is at least one kernel function for which inhomogeneous random graphs manifest clustering. However, the proof techniques we use to prove clustering in $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ rely heavily on the linearity of the inner product. Thus it is not clear whether there is a large class of kernel functions for which inhomogeneous random graphs will exhibit clustering or if such a class exists what should be the natural description of the class.

Finally, we consider the results on the degree distribution of inhomogeneous random graphs and $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$. Instead of stating the full generality of the results of Bollobás et al. on the degree distribution for inhomogeneous random graphs, we will rather apply their techniques explicitly to $\mathcal{G}_n^{\langle\cdot,\cdot\rangle}(\mu, n)$, in order to give a flavor for these results. First note that since the expected degree of a vertex is constant, if k is $\omega(1)$ then for a vertex v, we have $\mathbb{P}(\deg(v) = k) \to 0$. Thus we may assume that k is constant. Now as has been noted elsewhere [59, 60], in $\mathcal{G}_n^{\langle\cdot,\cdot\rangle}(\mu, n)$

$$\mathbb{P}(\deg(v) = k) = \int \binom{n-1}{k} \left(\frac{\langle \mathbb{E}[\mu], X\rangle}{n}\right)^k \left(1 - \frac{\langle \mathbb{E}[\mu], X\rangle}{n}\right)^{n-1-k} d\mu(X).$$

But then since $\binom{n-1}{k} \left(\frac{\langle \mathbb{E}[\mu], X\rangle}{n}\right)^k \left(1 - \frac{\langle \mathbb{E}[\mu], X\rangle}{n}\right)^{n-1-k} \leq 1$, by dominated convergence

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(\deg(v) = k) &= \int \lim_{n\to\infty} \binom{n-1}{k} \left(\frac{\langle \mathbb{E}[\mu], X\rangle}{n}\right)^k \left(1 - \frac{\langle \mathbb{E}[\mu], X\rangle}{n}\right)^{n-1-k} d\mu(X) \\ &= \int \frac{1}{k!} \langle \mathbb{E}[\mu], X\rangle^k e^{-\langle \mathbb{E}[\mu], X\rangle} d\mu(X)\end{aligned}$$

Recall that if Z is distributed as a Poisson distribution with parameter $\lambda \geq 0$, denoted

$\mathcal{P}(\lambda)$, then $\mathbb{P}(Z = k) = \frac{e^{-\lambda}\lambda^k}{k!}$. Hence, the limiting distribution of $\deg(v)$ is

$$\int \mathcal{P}(\langle \mathbb{E}[\mu], X\rangle)\, d\mu(X),$$

a Poisson mixture distribution with parameter distributed as $\mu^{\langle\cdot,\cdot\rangle}$. We notice that the key observation in this technique is the observation that $\mathcal{B}\big(n, \frac{\lambda}{n}\big) \to \mathcal{P}(\lambda)$ as $n \to \infty$. However, it is readily apparent that such a technique will not work when the sparsification function $g(n)$ is $o(n)$. Thus we need the more careful analysis of $\mathcal{B}(n, P)$ where P is a random variable. This analysis pays dividends not where k is $\Theta\Big(\frac{n}{g(n)}\Big)$ but outside that range where Theorem 2.3 describes the manner in which the asymptotic behavior depends on n. We note as well that the methodology of Theorem 2.3 should generalize to other "well-behaved" kernel functions, whereas it is hard to see the methodology of [13] generalizing to handle sparsification functions that are $o(n)$.

Although the work in [13] and this thesis both deal with $\mathcal{G}_g^{\kappa}(\mu, n)$ in a significant amount of generality, it is clear from this analysis that neither of the methodologies applied in either work will suffice to resolve the full generality of $\mathcal{G}_g^{\kappa}(\mu, n)$. But rather the behavior will have to be analyzed by a clever combination of the two sets of methodologies, perhaps coupled with some as yet unknown techniques.

CHAPTER III

DIRECTED RANDOM DOT PRODUCT GRAPHS

Although many of the phenomena that models for complex networks attempt to capture occur in directed networks, there has been a dearth of models that accurately capture this directed behavior in a natural and consistent manner. For instance, the natural way to adapt the preferential attachment model yields acyclic graphs in contrast to the readily observed short cycles in networks such as the World Wide Web. In this chapter we attempt to adapt $\mathcal{G}_g^{\langle \cdot,\cdot \rangle}(\mu, n)$ to directed networks and find that in many ways the hallmark properties of complex networks are maintained in this directed network. To this end we define directed random dot product graphs as follows.

Let $g(n) \geq 1$ and let $\mu \boxtimes \nu$ be a probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ where μ and ν are the marginal probability measures on the first and second d components, respectively. Thus a random variable distributed according to $\mu \boxtimes \nu$ will be represented by (X, Y), where X corresponds to the first d components and Y corresponds to the last d components. Note that it is not necessarily the case that $\mu \boxtimes \nu = \mu \times \nu$. Then the directed random dot product graph, denoted $\mathcal{G}_g^{\langle \cdot,\cdot \rangle}(\mu \boxtimes \nu, n)$, is the product measure $(\mu \boxtimes \nu)^n \times (\mathcal{U}[0,1])^{n^2-n}$ together with the map that takes elements of $\left(\mathbb{R}^d \times \mathbb{R}^d\right)^n \times [0,1]^{n^2-n}$, represented as $((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n), u_{1,2}, u_{1,3}, \ldots, u_{2,1}, u_{2,3}, \ldots, u_{n,n-1})$, to directed graphs on $[n]$ via the mapping where the arc (i, j) is present if and only if $\frac{\langle X_i, Y_j \rangle}{g(n)} \geq u_{i,j}$. Thus we may view $\mathcal{G}_g^{\langle \cdot,\cdot \rangle}(\mu \boxtimes \nu, n)$ as a probability measure on directed graphs on $[n]$. In addition, as in the definition of $\mathcal{G}_g^{\langle \cdot,\cdot \rangle}(\mu, n)$, we will assume that

$$((\mu \boxtimes \nu) \times (\mu \boxtimes \nu))(\{((X_u, Y_u), (X_v, Y_v)) \mid \langle X_u, Y_v \rangle \notin [0,1]\}) = 0,$$

we will call this the *inner product condition.* If in addition,

$$((\mu \boxtimes \nu) \times (\mu \boxtimes \nu))(\{((X_u, Y_u), (X_v, Y_v)) \mid \langle X_u, Y_v \rangle \notin (0,1]\}) = 0,$$

we will say that $\mu \boxtimes \nu$ satisfies the *strong inner product condition.*

Given that $\mu \boxtimes \nu$ satisfies the inner product condition, we may also view $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu \boxtimes \nu, n)$ in the following more constructive form:

- To each vertex v in a set of size n independently associate random variables (X_v, Y_v) distributed according $\mu \boxtimes \nu$
- Each arc (u, v) is present independently with probability $\frac{\langle X_u, Y_v \rangle}{g(n)}$.

We will also assume, similarly to $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$, that $\mu \boxtimes \nu$ is bounded, in the sense that $(\mu \boxtimes \nu)(\mathbb{B}(0;1) \times \mathbb{B}(0;1)) = 1$. In contrast to $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$, this assumption may limit the model in a non-trivial way. For instance, one can quickly construct a distribution where μ is concentrated at e_1 and the distribution determined by $\langle e_1, \nu \rangle$ lies entirely in $[0, 1]$ despite ν being unbounded. In this case, it is easy to see that there is a alternative bounded distribution $\mu \boxtimes \nu'$ formed by replacing ν by the projection of ν onto e_1. However, it is not obvious whether such modified distributions exist in general. But from a modelling/practical point of view, by the accepted assumption of bounded computational resources, we may as well assume that $(\mu \boxtimes \nu)(\mathbb{B}(0;1) \times \mathbb{B}(0;1)) = 1$.

Although this yields a model that is not, in the strict mathematical sense, a generalization of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ we will show in the remainder of this chapter that many of the same techniques used to analyze $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ in the previous chapter, will yield similar results for $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu \boxtimes \nu, n)$. In the first section, we generalize the notion of clustering to directed graphs and show that, depending on the behavior of $\mu \boxtimes \nu$, $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu \boxtimes \nu, n)$ exhibits various forms of clustering. Following that, we are able echo the results on the connectivity and diameter of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ by reducing the diameter and connectivity of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu \boxtimes \nu, n)$ to that of a directed Erdős-Rényí graph. Finally, in Section 3.3, we provide a characterization of the "logarithmic-histogram" of the in-degree and out-degrees of a vertex in $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu \boxtimes \nu, n)$.

3.1 Directed Clustering

In this section we consider the nature of the clustering that occurs in $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu \boxtimes \nu, n)$. For undirected graphs clustering is loosely defined as the presence of sets of vertices such that the induced subgraph is denser than would be "expected". Sometimes this is phrased in terms of a social interaction where it is more likely that the friend of a friend, is a friend of the original entity. In more mathematical terms, the presence of a path of length two increases the likelihood that the endpoints are connected by an edge. In Section 2.2, we showed that unless μ is constant, $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ exhibits this type of clustering in a probabilistic sense. However, for directed graphs it is not clear what is the appropriate generalization. Thus, instead of considering one particular generalization we consider all four possible orientations of the undirected structure. That is we consider

- $\mathbb{P}(u \to w \mid u \to v, v \to w)$,
- $\mathbb{P}(u \to w \mid u \to v, w \to v)$,
- $\mathbb{P}(u \to w \mid v \to u, v \to w)$, and
- $\mathbb{P}(u \to w \mid v \to u, w \to v)$.

In order to characterize the behavior of these probabilities we will find the following convexity lemma useful.

Lemma 3.1. *Let Ω be a real inner product space and let $a, b \in \Omega$. Let $D \subseteq \Omega$ be a region such that for all $x \in D$, $\langle a, x\rangle \in [0,1]$ and $\langle b, x\rangle \in [0,1]$. Then $u\colon D \longrightarrow \mathbb{R}$ defined by $x \longmapsto \langle a, x\rangle \langle x, b\rangle$ is a convex function of x.*

Proof. Let $F : (0,1) \times (0,1) \longrightarrow \mathbb{R}$ be defined by $(x, y) \longmapsto xy$. Then

$$\nabla^2 F = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

This matrix, although not positive semi-definite, is positive semi-definite over $[0,1] \times [0,1]$, and hence $F(x, y)$ is convex over its domain [8]. Since $\langle a, x\rangle$ is a real inner product, for any $\lambda \in [0,1]$ and $x, y \in D$, $\langle a, \lambda x + (1-\lambda)y\rangle = \lambda \langle a, x\rangle + (1-\lambda) \langle a, y\rangle$. Thus $\langle a, x\rangle$ is a convex function in x and similarly for $\langle b, x\rangle$. Thus $u(x) = F(\langle a, x\rangle, \langle b, x\rangle)$ is the composition of convex functions and is thus convex. □

Theorem 3.1. *Let u, v, w be vertices in $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu \boxtimes \nu, n)$. Then $\mathbb{P}(u \to w \mid u \to v, v \to w) \geq \mathbb{P}(u \to w)$. Further, if $\mu \boxtimes \nu = \mu \times \nu$, then*

- $\mathbb{P}(u \to w \mid u \to v, w \to v) \geq \mathbb{P}(u \to w)$,
- $\mathbb{P}(u \to w \mid v \to u, v \to w) \geq \mathbb{P}(u \to w)$, *and*
- $\mathbb{P}(u \to w \mid v \to u, w \to u) = \mathbb{P}(u \to w)$.

Proof. First we note that by Lemma 3.1 and Jensen's Inequality, for any two vectors a and b, we have

$$\int \langle a, X\rangle \langle b, X\rangle \, d(\mu \boxtimes \nu)(X, Y) \geq \langle a, \mathbb{E}[\mu]\rangle \langle b, \mathbb{E}[\mu]\rangle \quad \text{and}$$
$$\int \langle a, Y\rangle \langle b, Y\rangle \, d(\mu \boxtimes \nu)(X, Y) \geq \langle a, \mathbb{E}[\nu]\rangle \langle b, \mathbb{E}[\nu]\rangle .$$

Applying this, we then have that

$$\begin{aligned} &g(n)^3 \mathbb{P}(u \to w, u \to v, v \to w) \\ &\quad = \iiint \langle X_u, Y_w\rangle \langle X_u, Y_v\rangle \langle X_v, Y_w\rangle \, d\mu \boxtimes \nu(X_u, Y_u) d\mu \boxtimes \nu(X_v, Y_v) d\mu \boxtimes \nu(X_w, Y_w) \\ &\quad \geq \iint \langle \mathbb{E}[\mu], Y_w\rangle \langle \mathbb{E}[\mu], Y_v\rangle \langle X_v, Y_w\rangle \, d\mu \boxtimes \nu(X_v, Y_v) d\mu \boxtimes \nu(X_w, Y_w) \\ &\quad \geq \int \langle \mathbb{E}[\mu], \mathbb{E}[\nu]\rangle \langle \mathbb{E}[\mu], Y_v\rangle \langle X_v, \mathbb{E}[\nu]\rangle \, d\mu \boxtimes \nu(X_v, Y_v) \\ &\quad = g(n)^3 \mathbb{P}(u \to w) \, \mathbb{P}(u \to v, v \to w) . \end{aligned}$$

Thus $\mathbb{P}(u \to w \mid u \to v, v \to w) \geq \mathbb{P}(u \to w)$.

Now if $\mu \boxtimes \nu = \mu \times \nu$, then

$$\begin{aligned} &g(n)^3 \mathbb{P}(u \to w, u \to v, w \to v) \\ &\quad = \iiint \langle X_u, Y_w\rangle \langle X_u, Y_v\rangle \langle X_w, Y_v\rangle \, d\mu \boxtimes \nu(X_u, Y_u) d\mu \boxtimes \nu(X_v, Y_v) d\mu \boxtimes \nu(X_w, Y_w) \\ &\quad \geq \iint \langle \mathbb{E}[\mu], Y_w\rangle \langle Y_v, \mathbb{E}[\mu]\rangle \langle X_w, Y_v\rangle \, d\mu \boxtimes \nu(X_v, Y_v) d\mu \boxtimes \nu(X_w, Y_w) \\ &\quad = \int \langle \mathbb{E}[\mu], \mathbb{E}[\nu]\rangle \langle \mathbb{E}[\mu], Y_v\rangle \langle Y_v, \mathbb{E}[\mu]\rangle \, d\mu \boxtimes \nu(X_v, Y_v) \\ &\quad = g(n)^3 \mathbb{P}(u \to w) \, \mathbb{P}(u \to v, w \to v) . \end{aligned}$$

Similarly,

$$
\begin{aligned}
&g(n)^3\mathbb{P}(u \to w, v \to u, v \to w) \\
&\quad = \iiint \langle X_u, Y_w\rangle \langle X_v, Y_u\rangle \langle X_v, Y_w\rangle \, d\mu \boxtimes \nu(X_u, Y_u) d\mu \boxtimes \nu(X_v, Y_v) d\mu \boxtimes \nu(X_w, Y_w) \\
&\quad \geq \iint \langle X_u, \mathbb{E}[\nu]\rangle \langle X_v, Y_u\rangle \langle X_v, \mathbb{E}[\nu]\rangle \, d\mu \boxtimes \nu(X_u, Y_u) d\mu \boxtimes \nu(X_w, Y_w) \\
&\quad = \int \langle \mathbb{E}[\mu], \mathbb{E}[\nu]\rangle \langle \mathbb{E}[\mu], X_v\rangle \langle X_v, \mathbb{E}[\nu]\rangle \, d\mu \boxtimes \nu(X_v, Y_v) \\
&\quad = g(n)^3\mathbb{P}(u \to w)\, \mathbb{P}(v \to u, v \to w)\,.
\end{aligned}
$$

Finally

$$
\begin{aligned}
&g(n)^3\mathbb{P}(u \to w, w \to v, v \to u) \\
&\quad = \iiint \langle X_u, Y_w\rangle \langle X_w, Y_v\rangle \langle X_v, Y_w\rangle \, d\mu \boxtimes \nu(X_u, Y_u) d\mu \boxtimes \nu(X_v, Y_v) d\mu \boxtimes \nu(X_w, Y_w) \\
&\quad = \langle \mathbb{E}[\mu], \mathbb{E}[\nu]\rangle^3 \\
&\quad = g(n)^3\mathbb{P}(u \to w)\, \mathbb{P}(w \to v, v \to u)\,.
\end{aligned}
$$

Thus we have that if $\mu \boxtimes \nu = \mu \times \nu$, then

- $\mathbb{P}(u \to w \mid u \to v, w \to v) \geq \mathbb{P}(u \to w)$,
- $\mathbb{P}(u \to w \mid v \to u, v \to w) \geq \mathbb{P}(u \to w)$, and
- $\mathbb{P}(u \to w \mid v \to u, w \to u) = \mathbb{P}(u \to w)$.

□

3.2 Strong Connectivity and Directed Diameter

In this section we consider the connectivity properties of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu \boxtimes \nu, n)$. Although with directed graphs there are several different viewpoints from which to view the connectivity of the network, we choose the restrict ourselves to the most stringent of these, strong connectivity. However, our methods will easily extend to the consideration of weaker definitions of connectivity such as the connectivity of the underlying undirected network. With this in mind, we define the directed diameter ($\overrightarrow{\text{diam}}(G)$) to be the length of the maximum directed shortest path between two vertices in G. For graphs that are not strongly connected the directed diameter is infinite.

Theorem 3.2. *Suppose $r_\mu(n)$ and $r_\nu(n)$ are decreasing functions of n such that*

$$n\,(\mu \boxtimes \nu)\Big(\mathbb{B}(0; r_\mu(n)) \times \mathbb{R}^d \cup \mathbb{R}^d \times \mathbb{B}(0; r_\nu(n))\Big) \to 0,$$

$g(n)$ is monotonically increasing, and $\mu \boxtimes \nu$ satisfies the strong inner product condition. Then if there exists a function $d(n)$ such that for every $0 < c, C \leq 1$, there exists a c' such that

$$\mathbb{P}\left(\overrightarrow{\text{diam}}\left(\overrightarrow{\mathcal{G}}\left(cn, \frac{Cr_\mu(n) r_\nu(n)}{g(n)}\right)\right) \leq c' d(n)\right) = 1 - o(1)\,,$$

then with probability $1 - o(1)$ an arbitrarily large fraction of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu \boxtimes \nu, n)$ is strongly connected with diameter $\mathcal{O}(d(n))$.

Proof. First we observe that if $n\,(\mu \boxtimes \nu)\big(\mathbb{B}(0; r_\mu(n)) \times \mathbb{R}^d \cup \mathbb{R}^d \times \mathbb{B}(0; r_\nu(n))\big) \to 0$ then with probability $1 - o(1)$ there does not exist a vertex v such that $\|X_v\|_2 < r_\mu(n)$ or $\|Y_v\|_2 < r_\nu(n)$.

Let $\epsilon > 0$ and let $A(\delta) = \big\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \mid \langle x, y\rangle < \delta\big\}$. Then we observe that, since if (X, Y) is distributed according to $\mu \boxtimes \nu$, then $\langle X, Y\rangle > 0$ with probability 1, there exists a $\delta' > 0$ such that $(\mu \boxtimes \nu)(A(\delta')) < \frac{\epsilon}{2}$. But then by Chernoff Bound (C5) with probability at least $1 - e^{-\Theta(n)}$ there are at most ϵn vertices in $A(\delta')$. We thus restrict our attention to the vertices outside of $A(\delta')$. With this in mind define the compact set $K = \texttt{closure}\big(\mathbb{S}^d \times \mathbb{S}^d - A(\delta')\big)$, where $\mathbb{S}^d$ is the unit d-dimensional sphere.

Now for each $(x, y) \in \mathbb{S}^d \times \mathbb{S}^d$ define $\epsilon_{x,y} > 0$ such that if $x' \in \mathbb{S}^d$ with $\langle x', x\rangle > 1 - \epsilon_{x,y}$ and $y' \in \mathbb{S}^d$ with $\langle y', y\rangle > 1 - \epsilon_{x,y}$, then $\langle x', y'\rangle \geq \frac{\langle x,y\rangle}{2}$. Note that such an $\epsilon_{x,y}$ exists by the continuity of the inner product. Consider the open cover $\{\mathcal{C}(x, \epsilon_{xy}) \times \mathcal{C}(y, \epsilon_{xy})\}_{(x,y)\in K}$ of the set K. Now there is some finite subcover of K. Let $\{\mathcal{C}(x_i, \epsilon_{x_i y_i}) \times \mathcal{C}(y_i, \epsilon_{x_i y_i})\}_i$ be the sets in the finite subcover with positive measure with respect to $\mu \boxtimes \nu$. For notational convenience define $\mathcal{R}_i = \mathcal{C}(x_i, \epsilon_{xy}) \times \mathcal{C}(y_i, \epsilon_{xy})$ and $c = \frac{1}{2}\min_i \{(\mu \boxtimes \nu)(\mathcal{R}_i)\} > 0$. Then letting n_i be the number of vertices in region $\mathcal{R}_i$ we have that by Chernoff Bound (C5) $n_i \geq cn$ for all i with probability at least $1 - e^{-\Omega(n)}$. But then within the region R_i the probability that there is an arc from v to u is at least $\frac{\langle x_i, y_i\rangle r_\mu(\frac{n_i}{c}) r_\nu(\frac{n_i}{c})}{2g(\frac{n_i}{c})}$. This implies that for any c',

$$\mathbb{P}\Big(\overrightarrow{\text{diam}}(\mathcal{R}_i) \leq c'_i d(n)\Big) \geq \mathbb{P}\left(\overrightarrow{\text{diam}}\left(\overrightarrow{\mathcal{G}}\left(n_i, \frac{\langle x_i, y_i\rangle\, r_\mu(\frac{n_i}{c}) r_\nu(\frac{n_i}{c})}{2g(\frac{n_i}{c})}\right)\right) \leq c'_i d(\frac{n_i}{c})\right)$$
$$= \mathbb{P}\left(\overrightarrow{\text{diam}}\left(\overrightarrow{\mathcal{G}}\left(cn, \frac{\langle x_i, y_i\rangle\, r_\mu(n) r_\nu(n)}{2g(n)}\right)\right) \leq c'_i d(n)\right)$$
$$= 1 - o(1)\,.$$

Thus with probability at least $1 - o(1)$ every region $\mathcal{R}_i$ is strongly connected with directed diameter $\mathcal{O}(d(n))$.

At this point it suffices to show that for every $i \neq j$ there is are vertices $v_i \in \mathcal{R}_i$ and $v_j \in \mathcal{R}_j$ such that (v_i, v_j) is an edge. But since $\mu \boxtimes \nu$ satisfies the strong inner product condition, there is an $\epsilon'_{ij} > 0$ together with subregions $\mathcal{R}'_i \subseteq \mathcal{R}_i$ and $\mathcal{R}'_j \subseteq \mathcal{R}_j$ with positive measure, such that for any $(x_i, y_i) \in \mathcal{R}'_i$ and $(x_j, y_j) \in \mathcal{R}'_j$, $\langle x_i, y_j\rangle \geq \epsilon_{ij}$. However, this implies that the expected number of arcs from $\mathcal{R}_i$ to $\mathcal{R}_j$ tends to infinity, and thus with probability at least $1 - o(1)$ there is at least one such arc. Then, since there is an arc between every pair of regions and every region is strongly connected with directed diameter $\mathcal{O}(d(n))$ the subgraph formed by vertices contained with $\bigcup_i \mathcal{R}_i$ is strongly connected with directed diameter $\mathcal{O}(d(n))$. But then since $(\mu \boxtimes \nu)(\bigcup_i \mathcal{R}_i) \geq 1 - \epsilon$ and since ϵ is arbitrary, this implies that an arbitrarily large fraction of the graph is strongly connected with directed diameter $\mathcal{O}(d(n))$. □

We note that this result seems weaker than the corresponding result for $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ in that we cannot characterize the diameter of the entire graph, but rather we characterize the diameter of an arbitrarily large fraction of the vertices. However, it is worth noting that the vertices excluded are precisely those vertices v such that $\langle X_v, Y_v\rangle$ is small (that is less than some small δ'). In particular, if self-loops were allowed, these would be the vertices that are unlikely to have a self-loop. Interpreting this from a modeling point of view, these are the entities in the network that wouldn't want to "communicate" with themselves. Thinking of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu \boxtimes \nu, n)$ as a model for a social network, it is reasonable to assume that every entity has a bounded amount of "self-loathing". This would have the effect of assuring that there is a sufficiently small δ such that $(\mu \boxtimes \nu)(A(\delta)) = 0$ and thus, in effect, Theorem 3.2

characterizes the diameter of the entire graph rather than an arbitrarily large fraction.

3.3 Degree Distribution

Based on the intuition developed in the proof of Theorem 2.3, it is clear that if $\mu\boxtimes\nu \neq \mu\times\nu$, then in-degree and out-degree of a vertex will be coupled. In fact, in contrast to the undirected case (and the case where $\mu\boxtimes\nu = \mu\times\nu$), the distribution of $\left(\deg^+(v)\,,\deg^-(v)\right)$ will depend not on one random variable, but rather the following three random variables:

$$X_v^T\mathbb{E}\left[\nu^T\mu\right]Y_v, \qquad \langle X_v, \mathbb{E}\left[\nu\right]\rangle\,, \quad \text{and} \quad \langle\mathbb{E}\left[\mu\right], Y_v\rangle\,.$$

In addition, these random variables are not pair-wise independent, let alone independent as a set of random variables, which further complicates the analysis. Thus, in Theorem 3.3, we limit our consideration to the marginal distributions for $\left(\deg^+(v)\,,\deg^-(v)\right)$.

Further extending our intuition from Theorem 2.3 it is natural to define the following probability measures:

- $\mu^{\langle\cdot,\mathbb{E}[\nu]\rangle}(S) = \int_{\langle X,\mathbb{E}[\nu]\rangle\in S} d(\mu\boxtimes\nu)(X,Y)$,
- $\nu^{\langle\cdot,\mathbb{E}[\mu]\rangle}(S) = \int_{\langle\mathbb{E}[\mu],Y\rangle\in S} d(\mu\boxtimes\nu)(X,Y)$,
- $\mu_g^{\langle\cdot,\mathbb{E}[\nu]\rangle}(S) = \int_{\frac{\langle X,\mathbb{E}[\nu]\rangle}{g(n)}\in S} d(\mu\boxtimes\nu)(X,Y)$, and
- $\nu_g^{\langle\cdot,\mathbb{E}[\mu]\rangle}(S) = \int_{\frac{\langle\mathbb{E}[\mu],Y\rangle}{g(n)}\in S} d(\mu\boxtimes\nu)(X,Y)$.

We these definitions in hand, we can characterize the distributions of $\deg^+(v)$ and $\deg^-(v)$ as follows.

Theorem 3.3. *If $\mu^{\langle\cdot,\mathbb{E}[\nu]\rangle}$ is a Borel measure such that there exists a c^- such that for every measurable set S, $\mu^{\langle\cdot,\mathbb{E}[\nu]\rangle}(S) \leq c^-\mathcal{L}(S)$ and $g(n)$ is $o(n)$, then for any vertex v in $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu\boxtimes\nu, n)$ and any fixed $0<\delta<1$,*

$$\mathbb{P}\left(\left|\deg^-(v) - k\right| \leq \delta k\right) = \mu^{\langle\cdot,\mathbb{E}[\nu]\rangle}\left(\left[\frac{g(n)(1-\delta)k}{n-1}, \frac{g(n)(1+\delta)k}{n-1}\right]\right) + o(1)\,.$$

Similarly if $\nu^{\langle\cdot,\mathbb{E}[\mu]\rangle}$ is a Borel-measure such that there is a c^+ such that for every measurable set S, $\nu^{\langle\cdot,\mathbb{E}[\mu]\rangle}(S) \leq c^+\mathcal{L}(S)$ and $g(n)$ is $o(n)$, then for v in $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu\boxtimes\nu, n)$ and any *fixed* $0 < \delta < 1$,

$$\mathbb{P}\Big(\big|\deg^+(v) - k\big| \le \delta k\Big) = \nu^{\langle \cdot, \mathbb{E}[\mu]\rangle}\left(\left[\frac{g(n)(1-\delta)k}{n-1}, \frac{g(n)(1+\delta)k}{n-1}\right]\right) + o(1)\,.$$

Further, if k is $\omega(1)$, *then error term in either expression can be tightened to* $\mathcal{O}\left(\sqrt{\frac{\ln(k)}{k}}\right)$.

Proof. In a similar manner as Theorem 2.3, we have

$$\begin{aligned}\mathbb{P}\big(\deg^-(v) = k\big) &= \int \sum_{\substack{S\subseteq V\\ |S|=k}} \prod_{s\in S} \frac{\langle X_v, Y_s\rangle}{g(n)} \prod_{t\notin S}\left(1 - \frac{\langle X_v, Y_t\rangle}{g(n)}\right) \prod_{i\in V} d(\mu \boxtimes \nu)(X_i, Y_i)\\ &= \int \binom{n-1}{k}\left(\frac{\langle X_v, \mathbb{E}[\nu]\rangle}{g(n)}\right)^k \left(1 - \frac{\langle X_v, \mathbb{E}[\nu]\rangle}{g(n)}\right)^{n-1-k} d(\mu\boxtimes\nu)(X_v, Y_v)\\ &= \int \mathcal{B}\left(n-1, \frac{\langle X_v, \mathbb{E}[\nu]\rangle}{g(n)}\right) d\mu\boxtimes\nu(X_v, Y_v).\end{aligned}$$

Similarly

$$\mathbb{P}\big(\deg^+(v) = k\big) = \int \mathcal{B}\left(n-1, \frac{\langle \mathbb{E}[\mu], Y_v\rangle}{g(n)}\right) d\mu\boxtimes\nu(X_v, Y_V).$$

We now note that there are orthonormal matrices Q_μ and Q_ν such that $Q_\mu\mathbb{E}[\mu] = e_1 = Q_\nu\mathbb{E}[\nu]$. Let (X, Y) be distributed according to $\mu\boxtimes\nu$ and define Z_μ to be the random variable $\sqrt{\frac{\langle\mathbb{E}[\nu],\mathbb{E}[\nu]\rangle}{\langle\mathbb{E}[\mu],\mathbb{E}[\nu]\rangle}}\langle Q_\nu X, e_1\rangle$ and define Z_ν to be $\sqrt{\frac{\langle\mathbb{E}[\mu],\mathbb{E}[\mu]\rangle}{\langle\mathbb{E}[\mu],\mathbb{E}[\nu]\rangle}}\langle e_1, Q_\mu Y\rangle$. Then $\langle Z_\mu, \mathbb{E}[Z_\mu]\rangle = \langle Q_\nu X, e_1\rangle = \langle X, \mathbb{E}[\nu]\rangle$ and $\langle Z_\nu, \mathbb{E}[Z_\nu]\rangle = \langle e_1, Q_\mu Y\rangle = \langle \mathbb{E}[\mu], Y\rangle$. But this implies that there are probability measures μ' and ν' such that the probability that $\deg^-(v) = k$ (respectively, $\deg^+(v) = k$) in $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu\boxtimes\nu, n)$ is the same as the probability $\deg(v) = k$ in $\mathcal{G}(\mu', n)$ (respectively, $\mathcal{G}(\nu', n)$). Thus by Theorem 2.3, the desired result follows. □

CHAPTER IV

FURTHER STRUCTURAL AND ALGORITHMIC ASPECTS

Although much of the study of complex networks and their models has focused on the "small world" properties of diameter, clustering and degree distribution, recently, in an attempt to differentiate among the numerous potential models, there has been an increasing emphasis on secondary aspects of complex networks in addition to the small world properties. In this chapter we focus on two of those additional properties, constant conductance/spectral gap and assortativity. We show in Section 4.1 that for a large class of μ that are in a certain sense "near" the expected degree sequence model and for a large range of $g(n)$, the random graph $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ has constant conductance and spectral graph. Finally, in Section 4.2 we present a general structural interpretation of assortativity and show that although $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ exhibits positive assortativity and in the limit the assortativity is not much greater than that in Erdős-Rényí graph models.

4.1 Conductance and Spectral Properties

In the early 90's it was noted that to due the extraordinarily rapid growth of the Internet it would be unsurprising, in fact maybe even expected, for the underlying infrastructure of the Internet to collapse under the increased load within a "few" years [1]. As we observe today, the Internet never turned into a slag pile of melted routers and switches. In fact, it has continued to thrive and effectively absorb an ever increasing load. Despite relatively small increases in overall Internet infrastructure and local improvements in traffic routing, it appears that the Internet will continue to thrive indefinitely. It has been observed [32] that the ability of the Internet to handle this ever increasing load can be explained by the conductance properties of the underlying network. In fact, approximation algorithms for multi-commodity flow on graphs with good conductance and spectral properties yield near optimal congestion [55]. One way in which the conductance and spectral properties of a graph manifest is through the rapid mixing of a uniform random walk on the graph. That is, consider the rate at which the Markov chain with transition matrix AD^{-1} converges to the limiting distribution, where A is the adjacency matrix and D is the diagonal matrix of degrees. Now since AD^{-1} is the matrix for a Markov chain the maximum modulus occurs at the eigenvalue one. Thus letting $1 = \lambda_1, \lambda_2, \lambda_3, \cdots \lambda_k$ be the eigenvalues and $v_1, v_2, \cdots, v_k$ be the corresponding eigenvectors with norm one, we have that for any vector v, $\left(AD^{-1}\right)^t v = \sum_{i=1}^{k} \lambda_i^t \langle v_i, v \rangle v_i$. Thus if $\max_{i \neq 1} |\lambda_i| << 1$, then the resulting Markov Chain will mix rapidly. Hence, in order to show that a class of graphs has good conductance and spectral properties, and thus good algorithmic properties, it suffices to show that $\max_{i \neq 1} |\lambda_i|$ does not tend to one too rapidly with the number of vertices.

In addition to the study of the spectrum of AD^{-1} there is a more structural method to determine the conductance and spectral properties of a graph. The essences of this structural methodology is to measure the "worst" cut in the underlying graph with respect to the random walk and the limiting distribution. We notice that the vector $D\mathbb{1}$ has eigenvalue one with respect to AD^{-1} and thus the limiting distribution on a vertex is proportional to the degree of that vertex. This observation leads to the following natural quantity for a graph G,

$$\Phi(G) = \min_{S \subset V(G)} \frac{C(S, \overline{S})}{\min\left\{\mathrm{Vol}(S), \mathrm{Vol}\left(\overline{S}\right)\right\}},$$

where $C(S, \overline{S}) = \left|\left\{\{u, v\} \in E(G) \mid u \in S, v \in \overline{S}\right\}\right|$ and $\mathrm{Vol}(S) = \sum_{v \in S} \deg(v)$. The quantity $\Phi(G)$ is known as the conductance of the graph G. Now, intuitively, it would seem that lower conductance implies and small spectral gap. In fact, this intuition holds true as seen by the well known inequality [53, 23]

$$1 - 2\Phi(G) \leq \lambda_2 \leq 1 - \frac{\Phi(G)^2}{2}.$$

Progress has been made in recent years on analyzing the conductance/spectral gap of various models for complex networks using both spectral and combinatorial analysis [22, 32, 46]. It is with these successes in mind that we consider the following results towards characterizing the behavior of $\mathcal{G}_g^{\langle \cdot, \cdot \rangle}(\mu, n)$. In Section 4.1.1 we generalize the arguments of Chung, Lu, and Vu [22] on the spectral gap on the expected degree sequence model to a larger class of graphs generated with independent edges. We exhibit the power of this generalization in Section 4.1.1.2 by analyzing the spectral gap of the Stochastic Kronecker graph. Then, in Section 4.1.2, inspired by the work of Flaxman, Frieze, and Vera on geometric preferential attachment graphs [28, 29] and the importance of the underlying geometry of μ, we consider the behavior of spacial (geometric) cuts in $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$.

4.1.1 Spectral Analysis

In this section we consider spectral gap of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ directly via the analysis of the spectrum of an appropriate matrix. In particular, suppose that D is a diagonal matrix of the degrees of a graph G and A is the adjacency matrix. We then consider the spectrum of AD^{-1}, and in particular the second largest eigenvalue. If the second largest eigenvalue of AD^{-1} is bounded away from 1 (or approaches 1 sufficiently slowly) then spectral gap is sufficiently large to yield good algorithmic results. In order to more easily analyze the second eigenvalue of AD^{-1} we make the following standard observations.

Suppose then that v is an eigenvector of AD^{-1} with eigenvalue λ. Let $\hat{v} = D^{-\frac{1}{2}}v$ and consider

$$D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\hat{v} = D^{-\frac{1}{2}}AD^{-1}v = \lambda D^{-\frac{1}{2}}v = \lambda\hat{v}.$$

Thus $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ has the same spectrum as AD^{-1}. In addition, $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ is a symmetric matrix, which in general, leads to a more accessible spectrum. However, we are interested in the second largest eigenvalue of the matrix, so instead of studying the spectrum of $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ we concern ourselves with the spectral norm of $D^{-\frac{1}{2}}AD^{-\frac{1}{2}} - P_{v'}$, where $P_{v'}$ is a projection matrix onto a one-dimensional subspace of the first eigenspace. Letting $v' = D^{\frac{1}{2}}\mathbb{1}$, we have that

$$D^{-\frac{1}{2}}AD^{-\frac{1}{2}}v' = D^{-\frac{1}{2}}A\mathbb{1} = D^{-\frac{1}{2}}D\mathbb{1} = D^{\frac{1}{2}}\mathbb{1} = v'.$$

Thus v' lies within the first eigenspace of $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$. In particular, we consider the case where

$$P_{v'} = \frac{\langle v', v'\rangle}{\|v'\|_2^2} = \frac{1}{\sum_k \deg(k)} D^{\frac{1}{2}}KD^{\frac{1}{2}},$$

where K is the n-by-n matrix of all ones. Now define $M = D^{-\frac{1}{2}}AD^{-\frac{1}{2}} - \frac{1}{\sum_k \deg(k)} D^{\frac{1}{2}}KD^{\frac{1}{2}}$ and thus we concern ourselves with the spectral norm of M, that is the maximum modulus of the eigenvectors, which we will denote $\|M\|$.

4.1.1.1 Spectral Analysis of Non-Homogeneous Random Graphs

We will consider first a generalization of the methods used by Chung, Lu, and Vu [22] in determining the spectrum of the expected degree sequence model to general non-homogeneous random graphs. Later in this section, we will apply this result both to the Stochastic Kronecker model and $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$. In order to facilitate the presentation of the results, we define w_i to be the expected degree of vertex i, that is $e_i^T P\mathbb{1}$ where e_i is the i^{th} standard basis vector and $\mathbb{1}$ is the matrix of all ones. Further we let $w_{\min}$ be the minimum of the expected degrees and let W be the diagonal matrix of expected degrees so that $We_i = w_i$. We also note that in contrast to our earlier work with $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$, and following with the methodology of Chung, Lu, and Vu, we will allow self-loops. That is $p_{i,i}$ need not be zero. With this in mind, we have the following result, which bounds the spectral norm of M by the spectral norm of a deterministic matrix, T, where $t_{ij} = \frac{p_{ij}}{\sqrt{w_i w_j}} - \frac{\sqrt{w_i w_j}}{\sum_k w_k}$. Note that in the work of Chung et al. the matrix T is identically 0 and the entirety of the work of the following theorem is in adapting the proof of [22] to account for nonzero T.

Theorem 4.1. *The second largest eigenvalue of the normalized adjacency matrix, that is, $I - \mathcal{L}$, where $\mathcal{L}$ is the normalized Laplacian, of a random graph generated according to a symmetric matrix P with minimum expected degree at least $\log^2(n)$, $n \left\| W^{-\frac{1}{2}} T W^{-\frac{1}{2}} \right\|_\infty$ being $o(1)$, and $n w_{min}^2 \left(\frac{1}{\sum_k w_k} + \left\| W^{-\frac{1}{2}} T W^{\frac{-1}{2}} \right\|_\infty \right) \geq \log^6(n)$, is at most $\frac{5+o(1)}{\sqrt{w_{min}}} + (1 + o(1)) \|T\| + o(1)$, where T is some deterministic matrix depending only on P.*

Before proving this theorem it is useful to recall the following fact. For any matrix, A,

$$\|A\| = \max_{\|x\|_2 = 1} \left| x^T A x \right| .$$

Proof. Let A be the adjacency matrix of the graph, and let d_i be the degree of the vertex i. Then if K is the matrix of ones and D is the diagonal matrix of the degrees, we will consider $M = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} - \frac{1}{\sum_i d_i} D^{\frac{1}{2}} K D^{\frac{1}{2}}$. Note that the largest eigenvalue of M is the second largest eigenvalue of the normalized adjacency matrix. We will proceed by decomposing M into the sum of several matrices with less variability than M. To this end we will define w_i as the expected degree of vertex i, that is $w_i = e_i^T P\mathbb{1}$ and W as the diagonal matrix of expected degrees. Then define the following matrices, where C, S, and R are the same matrices as in [22], H is a multiplicative "error matrix" implicitly present in [22], and the matrix T encapsulates the difference between the expected degree sequence model and a general inhomogeneous random graph model.

$$C = W^{-\frac{1}{2}}(A-P)W^{-\frac{1}{2}} \qquad c_{ij} = \frac{a_{ij}-p_{ij}}{\sqrt{w_i w_j}}$$

$$T = W^{-\frac{1}{2}}PW^{-\frac{1}{2}} - \frac{1}{\sum_k w_k}W^{\frac{1}{2}}KW^{\frac{1}{2}} \qquad t_{ij} = \frac{p_{ij}}{\sqrt{w_i w_j}} - \frac{\sqrt{w_i w_j}}{\sum_k w_k}$$

$$S = \left(\frac{1}{\sum_k w_k} - \frac{1}{\sum_k d_k}\right) D^{\frac{1}{2}}KD^{\frac{1}{2}} \qquad s_{ij} = \frac{\sqrt{d_i d_j}}{\sum_k w_k} - \frac{\sqrt{d_i d_j}}{\sum_k d_k}$$

$$R = \frac{1}{\sum_k w_k} D^{-\frac{1}{2}}\left(WKW - DKD\right) D^{-\frac{1}{2}} \qquad r_{ij} = \frac{w_i w_j}{\sqrt{d_i d_j}\sum_k w_k} - \frac{\sqrt{d_i d_j}}{\sum_k w_k}$$

$$H = W^{\frac{1}{2}}D^{-\frac{1}{2}}KD^{-\frac{1}{2}}W^{\frac{1}{2}} - K \qquad h_{ij} = \frac{\sqrt{w_i w_j}}{\sqrt{d_i d_j}} - 1.$$

It is then a simple exercise to see that $m_{ij} = (h_{ij}+1)(c_{ij}+t_{ij}) + s_{ij} + r_{ij}$, and thus $M = C + H \bullet C + T + H \bullet T + S + R$ where $\bullet$ is the Hadamard product. Thus $\|M\| \leq \|C\| + \|H \bullet C\| + \|T\| + \|H \bullet T\| + \|S\| + \|R\|$. We will prove this result via a series of four claims adapting the result of Chung et al. [22], where Claim 1 is implicit in their work, Claim 2 is directly from their work, and Claims 3 and 4 adapt the work of Chung, Lu, and Vu to the more general case in order to account for the matrix T.

Claim 1 *For any matrix N, $\|H \bullet N\|$ is $o(\|N\|)$ with probability at least $1 - e^{-\Theta\left(\log^2(n)\right)}$.*

Note that by Chernoff Bound (C1) $\mathbb{P}(|d_i - w_i| > \epsilon w_i) \leq e^{-\frac{3\epsilon^2 w_i}{6+2\epsilon}}$ for any fixed $\epsilon > 0$. Thus the probability that the degree of any vertex fails to be within ϵ of the expected degree is at most $ne^{-\frac{3\epsilon^2 w_i}{3+\epsilon}} \leq e^{-\Theta\left(\log^2(n)\right)}$. Now consider

$$\begin{aligned}
\|H \bullet N\| &= \max_{\|y\|_2=1} y^T(H \bullet N)y \\
&= \max_{\|y\|_2=1} \sum_{ij} y_i y_j n_{ij} \left(\frac{\sqrt{w_i w_j} - \sqrt{d_i d_j}}{\sqrt{d_i d_j}} \right) \\
&= \max_{\|y\|_2=1} \sum_{ij} y_i y_j n_{ij} \left(\frac{\sqrt{w_i w_j} - \sqrt{d_i d_j}}{\sqrt{d_i d_j}} \right) \\
&= \max_{\|y\|_2=1} \sum_{ij} y_i y_j n_{ij} \left(\frac{\sqrt{d_i}(\sqrt{w_j} - \sqrt{d_j}) + (\sqrt{w_i} - \sqrt{d_i})\sqrt{w_j}}{\sqrt{d_i d_j}} \right) \\
&= \max_{\|y\|_2=1} \sum_{ij} y_i n_{ij} \frac{\sqrt{w_j} - \sqrt{d_j}}{\sqrt{d_j}} y_j + \frac{\sqrt{w_i} - \sqrt{d_i}}{\sqrt{d_i}} y_i n_{ij} \sqrt{\frac{w_j}{d_j}} y_j
\end{aligned}$$

Now letting $y_i' = y_i \frac{\sqrt{w_i}-\sqrt{d_i}}{\sqrt{d_i}}$ and $y_j'' = y_j \sqrt{\frac{w_j}{d_j}}$, this yields

$$\|H \bullet N\| = \max_{\|y\|_2=1} \left\langle y, Ny' \right\rangle + \left\langle y', Ny'' \right\rangle.$$

Thus $\|H \bullet N\| \leq \|y'\|_2 \left(\|y\|_2 + \|y''\|_2\right) \|N\|$. By noting that $\|y'\|_2^2 \leq \frac{2-\epsilon}{1-\epsilon} - \frac{2}{\sqrt{1+\epsilon}} \to 0$ and $\|y''\|_2^2 \leq \frac{1}{1-\epsilon} \to 1$ as $\epsilon \to 0$, we have that $\|H \bullet N\|$ is $o(\|N\|)$ with high probability.

Claim 2 $\|S\| \leq \frac{h(n)}{\sqrt{\sum_i w_i}}$ *with probability at least* $1 - e^{-\Theta\left(\frac{\sqrt{\sum_i w_i} h(n)^2}{\sqrt{\sum_i w_i} + h(n)}\right)}$, *where* $h(n)$ *is an arbitrary slow growing function of* n.

Again, we observe that by Chernoff Bound (C1), $\mathbb{P}\left(|\sum_i d_i - \sum_i w_i| > h(n)\sqrt{\sum_i w_i}\right) < e^{\frac{3\sqrt{\sum_i w_i} h(n)^2}{3\sqrt{\sum_i w_i} + h(n)}}$. Now we observe that

$$\begin{aligned}
\|S\| &\leq \max_{\|y\|_2=1} \sum_{ij} \left| y_i y_j \left(\frac{1}{\sum_i d_i} - \frac{1}{\sum_i w_i} \right) \sqrt{d_i d_j} \right| \\
&= \left| \frac{\sum_i w_i - \sum_i d_i}{\sum_i w_i \sum_i d_i} \right| \max_{\|y\|_2=1} \sum_{ij} \left| y_i \sqrt{d_i} \right| \left| y_j \sqrt{d_j} \right|.
\end{aligned}$$

Letting $y_i''' = \sqrt{d_i} y_i$, we have that

$$\|S\| \leq \left| \frac{\sum_i w_i - \sum_i d_i}{\sum_i w_i \sum_i d_i} \right| \max_{\|y\|_2=1} \left\|y'''\right\|_2^2 \leq \left| \frac{h(n)\sqrt{\sum_i w_i}}{\sum_i w_i \sum_i d_i} \right| \max_{\|y\|_2=1} \|y\|_2^2 \sum_i d_i = \frac{h(n)}{\sqrt{\sum_i w_i}}.$$

Claim 3 *If* w_{min} *is at least* $\log^2(n)$ *and* $n \left\| W^{-\frac{1}{2}} T W^{-\frac{1}{2}} \right\|_\infty$ *is* $o(1)$, *then* $\|R\| \leq 3\sqrt{\frac{n}{\sum_i w_i}}$ *with probability at least* $1 - \frac{6}{h^2(n)}$, *where* $h(n)$ *is an arbitrarily slow growing function.*

We begin by considering

$$\begin{aligned}
\|R\| &\le \max_{\|y\|_2=1} \sum_{ij} y_i y_j \left| \frac{w_i w_j - d_i d_j}{\sqrt{d_i d_j}} \right| \\
&= \max_{\|y\|_2=1} \sum_{ij} y_i y_j \left| \frac{d_i(w_j - d_j) + w_j(w_i - d_i)}{\sqrt{d_i d_j}} \right| \\
&= \max_{\|y\|_2=1} \sum_i y_i \sqrt{d_i} \sum_j \left| \frac{y_j(w_j - d_j)}{\sqrt{d_j}} \right| + \sum_j \frac{w_j}{\sqrt{d_j}} y_j \sum_i \left| \frac{y_i(w_i - d_i)}{\sqrt{d_i}} \right| \\
&\le \max_{\|y\|_2=1} \left(\|y\|_2 \sqrt{\sum_k d_k} + \|y\|_2 \sqrt{\sum_k d_k} \right) \|y\|_2 \sqrt{\sum_i \frac{(w_i - d_i)^2}{d_i}} \\
&= \left(\sqrt{\sum_k d_k} + \sqrt{\sum_k w_k} \right) \sqrt{\sum_k \frac{(w_k - d_k)^2}{d_k}}.
\end{aligned}$$

But as in Claim 1, with high probability $|d_i - w_i| \le \epsilon w_i$ for any fixed $\epsilon > 0$, thus

$$\|R\| \le \frac{1 + \sqrt{1+\epsilon}}{\sqrt{1-\epsilon}} \sqrt{\sum_k w_k} \sqrt{\sum_k \frac{(w_k - d_k)^2}{w_k}} \tag{4.1}$$

Now let X_i be the random variable $(w_i - d_i)^2$ and let $X = \sum_i \frac{1}{w_i} X_i$. We wish to show that with high probability X is at most $(5 + o(1))n$. First observe that $X_i = (\sum_j a_{ij} - p_{ij})^2 = \sum_j (a_{ij} - p_{ij})^2$ since $\mathbb{E}[a_{ij} - p_{ij}] = 0$. Thus $\mathbb{E}[X_i] = \sum_j \mathbb{E}\left[(a_{ij} - p_{ij})^2\right] \le w_i$. Consider then

$$\begin{aligned}
\mathbb{E}\left[X_i^2\right] &= \mathbb{E}\left[\left(\sum_j a_{ij} - p_{ij} \right)^4 \right] \\
&= \sum_{j \ne k} \binom{4}{2} \mathbb{E}\left[(a_{ij} - p_{ij})^2 (a_{ik} - p_{ik})^2\right] + \sum_j \mathbb{E}\left[(a_{ij} - p_{ij})^4\right] \\
&= 6\mathbb{E}[X_i]^2 + \sum_j \mathbb{E}\left[(a_{ij} - p_{ij})^4\right] - \sum_j \mathbb{E}\left[(a_{ij} - p_{ij})^2\right]^2 \\
&\le w_i + 6\mathbb{E}[X_i]^2.
\end{aligned}$$

We now consider

$$\begin{aligned}
\mathbb{E}[X_i X_j] &= \mathbb{E}\left[\left(\sum_k a_{ik} - p_{ik} \right)^2 \left(\sum_\ell a_{j\ell} - p_{j\ell} \right)^2 \right] \\
&= \mathbb{E}[X_i]\,\mathbb{E}[X_j] + \mathbb{E}\left[(a_{ij} - p_{ij})^4\right] - \mathbb{E}\left[(a_{ij} - p_{ij})^2\right]^2 \\
&\le \mathbb{E}[X_i]\,\mathbb{E}[X_j] + p_{ij}.
\end{aligned}$$

Thus

$$\mathrm{Var}(X_i) \le w_i + 5\mathbb{E}[X_i]^2 \le w_i + 5w_i^2 \quad \text{and} \quad \mathrm{cov}(X_i, X_j) \le p_{ij}.$$

Further,

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_k \frac{1}{w_k^2}\mathrm{Var}(X_k) + 2\sum_{i<j}\frac{1}{w_i w_j}\mathrm{cov}(X_i X_j) \\
&= \sum_k (5 + \frac{1}{w_k}) + 2\sum_{i<j}\frac{p_{ij}}{w_i w_j} \\
&\le (5 + \frac{1}{w_{\min}})n + 2\sum_{i<j}\frac{t_{ij}}{\sqrt{w_i w_j}} + \frac{1}{\sum_k w_k} \\
&\le (5 + \frac{2}{w_{\min}})n + n^2\left\|W^{-\frac{1}{2}}TW^{-\frac{1}{2}}\right\|_\infty \\
&= \left(5 + o(1) + n\left\|W^{-\frac{1}{2}}TW^{\frac{-1}{2}}\right\|_\infty\right)n.
\end{aligned}$$

But since $n\left\|W^{-\frac{1}{2}}TW^{-\frac{1}{2}}\right\|_\infty$ is $o(1)$, $\mathrm{Var}(X) = (5 + o(1))n$. But then by Chebychev's inequality $\mathbb{P}(|X - \mathbb{E}[X]| \ge h(n)\sqrt{n}) \le \frac{(5+o(1))n}{nh^2(n)} = \frac{5+o(1)}{h^2(n)}$. Combining this with Equation 4.1, we have that if $h(n)$ is an arbitrarily slow growing function, then $\|R\| \le 3\sqrt{\frac{n}{\sum_k w_k}}$ with probability at least $1 - o(1)$.

Claim 4 *If* $nw_{min}^2\left(\frac{1}{\sum_i w_i} + \left\|W^{-\frac{1}{2}}TW^{-\frac{1}{2}}\right\|_\infty\right) \ge \log^6(n)$, *then*

$$\|C\| \le (1 + o(1))2\sqrt{\frac{n}{\sum_i w_i} + n\left\|W^{-\frac{1}{2}}TW^{-\frac{1}{2}}\right\|_\infty}$$

with probability at least $1 - o(1)$.

In order to bound $\|C\|$ we will use Wigner's high moment method in a similar manner to the argument of Chung, Lu, and Vu in [22]. Let $\mathcal{W}_t$ be the set of closed walks on K_n of length t. We note that $\mathtt{trace}(C^t) = \sum_{W\in W_t}\prod_{\{i,j\}\in W} c_{ij}^{|W\cap\{\{i,j\}\}|}$. Since $\mathbb{E}[c_{ij}] = 0$, we have that for any closed walk W which contains the edge i,j precisely once, $\mathbb{E}\left[\prod_{\{i,j\}\in W} c_{ij}^{|W\cup\{\{i,j\}\}|}\right] = 0$. We thus consider the set of closed walks of length $2k$ on K_n using ℓ distinct vertices. Denote the set of such walks as $W_{k,\ell}$. In order to bound $\mathbb{E}\left[\mathtt{trace}(C^{2k})\right]$, we will uniformly bound the contribution of each edge in a walk in $W_{k,\ell}$ based on its multiplicity. Note that

$$\begin{aligned}
\left|\mathbb{E}\left[c_{ij}^m\right]\right| &= \frac{(1-p_{ij})^m p_{ij} + p_{ij}^m(1-p_{ij})}{(\sqrt{w_i w_j})^m} \\
&\leq \frac{p_{ij}}{(w_i w_j)^{\frac{m}{2}}} \\
&\leq \frac{1}{w_{\min}^{m-2}} \frac{p_{ij}}{w_i w_j} \\
&= \frac{1}{w_{\min}^{m-2}} \left(\frac{1}{\sum_v w_v} + \frac{p_{ij}\sum_v w_v - w_i w_j}{w_i w_j \sum_v w_v} \right) \\
&= \frac{1}{w_{\min}^{m-2}} \left(\frac{1}{\sum_v w_v} + \frac{t_{ij}}{\sqrt{w_i w_j}} \right).
\end{aligned}$$

Thus, letting $\rho = \frac{1}{\sum_v w_v} + \frac{t_{ij}}{\sqrt{w_i w_j}}$ in order to echo the notation of Chung et al., the weight of a walk in $W_{k,\ell}$ is at most

$$\frac{\rho^\ell}{w_{\min}^{2k-2\ell}}. \tag{4.2}$$

In [31] Füredi and Komlós showed that

$$|W_{k,\ell}| \leq n(n-1)\cdots(n-l)\binom{2k}{2\ell}\binom{2\ell}{\ell}(\ell+1)^{4(k-\ell)-1} \leq n^{\ell+1}4^\ell\binom{2k}{2\ell}(\ell+1)^{4(k-l)}. \tag{4.3}$$

Combining (4.2) and (4.3), we have that

$$\begin{aligned}
\mathbb{E}\left[\texttt{trace}\left(C^{2k}\right)\right] &\leq \sum_{l=1}^{k} n \frac{(4\rho n)^\ell}{\left(w_{\min}^2(\ell+1)^4\right)^{k-\ell}} \binom{2k}{2\ell} \\
&= n\,(4\rho n)^k \left(1 + \sum_{\ell=1}^{k-1} \binom{2k}{2\ell} \left(\frac{4\rho n w_{\min}^2}{(\ell+1)^4}\right)^{\ell-k}\right) \\
&\leq n\,(4\rho n)^k \left(1 + \sum_{\ell=1}^{k-1} 2k^{2(k-\ell)} \left(\frac{4\rho n w_{\min}^2}{(\ell+1)^4}\right)^{\ell-k}\right) \\
&\leq n\,(4\rho n)^k \left(1 + \sum_{\ell=1}^{k-1} 2 \left(\frac{4\rho n w_{\min}^2}{k^6}\right)^{\ell-k}\right) \\
&= n\,(4\rho n)^k \left(1 + 2\sum_{\ell=1}^{k-1} \left(\frac{k^6}{4\rho n w_{\min}^2}\right)^{\ell}\right).
\end{aligned}$$

Observing that the summation is a geometric series, we have that

$$\mathbb{E}\left[\texttt{trace}\left(C^{2k}\right)\right] \leq n\,(4\rho n)^k \left(1 + 2\frac{k^6}{4n\rho w_{\min}^2 - k^6}\left(1 - \left(\frac{k^6}{4n\rho w_{\min}^2}\right)^{k-1}\right)\right).$$

Since,

$$n\rho w_{\min}^2 = n w_{\min}^2 \left(\frac{1}{\sum_i w_i} + \left\| W^{-\frac{1}{2}} T W^{-\frac{1}{2}} \right\|_\infty \right) \to \infty,$$

letting $k = \sqrt[6]{n\rho w_{\min}^2}$ yields that $\mathbb{E}\left[\mathtt{trace}\left(C^{2k}\right)\right] \leq 2n(2\sqrt{n\rho})^{2k}$. Thus, by using Markov's inequality,

$$\begin{aligned}
\mathbb{P}(\|C\| \geq (1+\epsilon)2\sqrt{n\rho}) &= \mathbb{P}\Big(\|C\|^{2k} \geq (1+\epsilon)^{2k} \left(2\sqrt{n\rho}\right)^{2k}\Big) \\
&\leq \mathbb{P}\Big(\mathtt{trace}\Big(C^{2k}\Big) \geq (1+\epsilon)^{2k} \left(2\sqrt{n\rho}\right)^{2k}\Big) \\
&\leq \frac{\mathbb{E}\left[\mathtt{trace}\left(C^{2k}\right)\right]}{(1+\epsilon)^{2k} \left(2\sqrt{n\rho}\right)^{2k}} \\
&\leq \frac{2n \left(2\sqrt{n\rho}\right)^{2k}}{(1+\epsilon)^{2k} \left(2\sqrt{n\rho}\right)^{2k}} \\
&= \frac{2n}{(1+\epsilon)^{2k}}
\end{aligned}$$

Since $k = \sqrt[6]{n\rho w_{\min}^2} \geq \log(n)$, there is some function $\epsilon(n)$ going to 0, such that $\frac{2n}{(1+\epsilon)^{2k}} \to 0$.

Combining Claims 1–4, we have that $\|M\| \leq (1+o(1))2\sqrt{n\rho} + (1+o(1))\|T\| + \frac{3+o(1)}{\sqrt{w_{\min}}}$. But

$$\sqrt{n\rho} = \sqrt{\frac{n}{\sum_i w_i} + n\left\| W^{-\frac{1}{2}} T W^{-\frac{1}{2}} \right\|_\infty} \leq \sqrt{\frac{n}{\sum_i w_i}} + \sqrt{n\left\| W^{-\frac{1}{2}} T W^{-\frac{1}{2}} T \right\|_\infty} \leq \frac{1}{\sqrt{w_{\min}}} + o(1).$$

Hence, $\|M\| \leq \frac{5+o(1)}{\sqrt{w_{\min}}} + (1+o(1))\|T\| + o(1)$. □

4.1.1.2 Spectrum of Stochastic Kronecker Graphs

As an aside to illustrate the power of Theorem 4.1, we will show that the second eigenvalue of the normalized adjacency matrix for the Stochastic Kronecker graph is at most $(1 + o(1))\frac{|\beta^2 - \alpha\gamma|}{(\alpha+\beta)(\beta+\gamma)}$. We recall that the probability of an edge is defined by the repeated Kronecker product of a small symmetric matrix where each entry is between zero and one. As a specific example we will consider the generating matrix

$$\begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix}$$

with $\gamma \leq \beta \leq \alpha$. This choice of restrictions on α, β, and γ is motivated by the observations of Leskovec and Faloutsos [41] that this choice of parameters results in the best approximation of real world networks. Since the generating matrix is 2×2, we may associated to each vertex a bit string of length $t = \log_2(n)$, where the i^{th} entry represents which row/column the vertex belongs to in the i^{th} Kronecker product. Thus the probability of a edge between u and v is $\alpha^{a_{uv}}\gamma^{g_{uv}}\beta^{t-a_{uv}-g_{uv}}$ where a_{uv} is the number of common zeros in the strings for u and v, and similarly g_{uv} is the number of common ones is the strings for u and v. From this it is easy to see that the expected degree of a vertex is u is $(\alpha+\beta)^{z_u}(\gamma+\beta)^{t-z_u}$ where z_u is the number of zeros in the string for u. Finally this gives $(\alpha+2\beta+\gamma)^t$ as the sum of the expected degrees.

Theorem 4.2. *If $\gamma \leq \beta \leq \alpha$ and $\frac{2(\beta+\gamma)^2}{\alpha+2\beta+\gamma} > 1$, the second eigenvalue for the normalized adjacency matrix for the Stochastic Kronecker graph is at most $(1+o(1))\frac{|\beta^2-\alpha\gamma|}{(\alpha+\beta)(\beta+\gamma)}$ with high probability.*

Proof. First note that $\frac{2(\beta+\gamma)^2}{\alpha+2\beta+\gamma} > 1$ implies that $\beta+\gamma > 1$. Since $w_{\min} = (\gamma+\beta)^t > t^2 = \log^2(n)$ and $\frac{nw_{\min}^2}{\sum_k w_k} = \frac{2^t(\beta+\gamma)^{2t}}{(\alpha+2\beta+\gamma)^t} > \log^6(n)$, by Theorem 4.1, we need only concern ourselves with the behavior of the matrices T and $W^{-\frac{1}{2}}TW^{-\frac{1}{2}}$, where $t_{ij} = \frac{p_{ij}}{\sqrt{w_iw_j}} - \frac{\sqrt{w_iw_j}}{\sum_k w_k}$, p_{ij} is the probability of the edge $\{i,j\}$, and w_k is the expected degree of vertex k. We observe that $\frac{p_{ij}}{\sqrt{w_iw_j}}$ is a Kronecker matrix with generating matrix

$$\begin{bmatrix} \frac{\alpha}{\alpha+\beta} & \frac{\beta}{\sqrt{(\alpha+\beta)(\beta+\gamma)}} \\ \frac{\beta}{\sqrt{(\alpha+\beta)(\beta+\gamma)}} & \frac{\gamma}{\beta+\gamma} \end{bmatrix}. \tag{4.4}$$

Further, $\frac{\sqrt{w_iw_j}}{\sum_k w_k}$ is a Kronecker matrix with generating matrix

$$\frac{1}{\alpha+2\beta+\gamma}\begin{bmatrix} \alpha+\beta & \sqrt{(\alpha+\beta)(\beta+\gamma)} \\ \sqrt{(\alpha+\beta)(\beta+\gamma)} & \beta+\gamma \end{bmatrix}. \tag{4.5}$$

It is well known that the eigenspaces and eigenvalues of a Kronecker product of matrices are completely determined by the eigenspaces of the multiplicands. But both (4.4) and (4.5) have the same eigenspaces, specifically those generated by

$$\begin{bmatrix} \sqrt{\alpha+\beta} \\ \sqrt{\beta+\gamma} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -\sqrt{\alpha+\beta} \\ \sqrt{\beta+\gamma} \end{bmatrix}.$$

For (4.4) the eigenvalues are 1 and $\frac{\alpha\gamma-\beta^2}{(\alpha+beta)(\beta+\gamma)}$, respectively, and for (4.5) the eigenvalues are 1 and 0, respectively. Thus $\|T\| = \frac{|\beta^2-\alpha\gamma|}{(\alpha+\beta)(\beta+\gamma)}$.

We now proceed to show that $n\left\|W^{-\frac{1}{2}}TW^{-\frac{1}{2}}\right\|_\infty \to 0$. First we note that $\frac{n}{\sum_k w_k} = \left(\frac{2}{\alpha+2\beta+\gamma}\right)^t \to 0$. Thus it suffices to consider the limiting behavior of $\frac{np_{ij}}{w_i w_j \sum_k w_k}$. For any particular i, j this quantity is

$$\left(\frac{2\beta}{(\alpha+\beta)(\beta+\gamma)(\alpha+2\beta+\gamma)}\right)^t \left(\frac{\alpha(\beta+\gamma)}{\beta(\alpha+\beta)}\right)^{a_{ij}} \left(\frac{\gamma(\alpha+\beta)}{\beta(\beta+\gamma)}\right)^{g_{ij}}. \tag{4.6}$$

Note that $0 \leq a_{ij} + g_{ij} \leq t$. In order to establish an upper bound the three cases that need to be considered are

1. $a_{ij} = 0, g_{ij} = 0$,

2. $a_{ij} = t, g_{ij} = 0$, and

3. $a_{ij} = 0, g_{ij} = t$.

In the first case $2\beta < \alpha + 2\beta + \gamma$ and so (4.6) approaches zero. For the second case (4.6) simplifies to $\frac{2\alpha}{(\alpha+\beta)^2(\alpha+2\beta+\gamma)}$. But $2 < \alpha+2\beta+\gamma$ and $\alpha < (\alpha+\beta)^2$ so this approaches zero as well. Finally, in the third case, (4.6) simplifies to $\frac{2\gamma}{(\beta+\gamma)^2(\alpha+2\beta+\gamma)}$ and similarly approaches 0. Thus since $\|T\|$ is constant, the second eigenvalue for the normalized adjacency matrix is at most $(1 + o(1))\,\|T\| = (1 + o(1))\frac{|\beta^2-\alpha\gamma|}{(\alpha+\beta)(\beta+\gamma)}$, which is asymptotically bounded away from 1. □

4.1.1.3 Spectral Analysis of Random Dot Product Graph

We consider in this section the implications of Theorem 4.1 on the spectral properties of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$. Note that in order to agree with the format of the theorem, we will consider a slight modification of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ where self-loops are allowed and occur with the natural probability. The driving force of the spectral theorem is the relationship between the expected degrees and the probability of an edge between two vertices. To that end we note that

$$p_{ij} = \frac{\langle X_i, X_j \rangle}{g(n)}$$
$$w_i = \frac{\left\langle X_i, \sum_j X_j \right\rangle}{g(n)}.$$

In order to more fully understand w_i and in particular $\sum_j X_j$ we note the following lemma.

Lemma 4.1. *Let $X_1, X_2, \ldots, X_n$ be random vectors in $\mathbb{R}^d$ distributed according to a probability measure μ with 2-norm at most 1 and let $X = \sum_j X_j$. Then $\mathbb{P}(\|X - \mathbb{E}[X]\|_\infty > \epsilon n) \leq de^{-\frac{\epsilon^2 n}{2+\frac{2\epsilon}{3}}}$.*

Proof. We proceed by considering each component individually. We note that since $\|X_j\|_2 \leq 1$, $|\langle X_j, e_i \rangle| \leq 1$ and further that $\mathbb{E}\left[\sum_j \langle X_j, e_i \rangle\right] = \left\langle \mathbb{E}\left[\sum_j X_j\right], e_i \right\rangle$. Now since each component is bounded in magnitude by 1, then $\mathrm{Var}(\langle X_j, e_i \rangle) \leq 1$ and so $\mathrm{Var}\left(\left\langle \sum_j X_j, e_i \right\rangle\right) \leq n$. Thus, by Chernoff Bound (C1), $\mathbb{P}\left(\left|\left\langle \sum_j X_j, e_i \right\rangle - \mathbb{E}\left[\left\langle \sum_j X_j, e_i \right\rangle\right]\right| > \epsilon n\right) \leq e^{-\frac{\epsilon^2 n}{2+\frac{2\epsilon}{3}}}$. The result follows by summing over all components. □

Thus for any ϵ such that $\frac{1}{\epsilon}$ is $o(\sqrt{n})$, with high probability $w_i = \frac{n}{g(n)} \langle X_i, \mathbb{E}[\mu] + \zeta \rangle$ where ζ is does not depend on i and $\|\zeta\|_2^2 \leq d\epsilon^2$. Thus $w_{\min} = \min_i \frac{n}{g(n)} \langle X_i, \mathbb{E}[\mu] + \zeta \rangle \geq \frac{n}{g(n)} \left(\min_i \langle X_i, \mathbb{E}[\mu] \rangle - \epsilon\sqrt{d}\right)$.

Now we turn to the consideration of the matrix $W^{-\frac{1}{2}} T W^{-\frac{1}{2}}$. First note that the i, j entry in this matrix is $\frac{p_{ij}}{w_i w_j} - \frac{1}{\sum_k w_k}$. We note that

$$\begin{aligned}
\sum_k w_k &= \sum_k \frac{n}{g(n)} \langle X_k, \mathbb{E}[\mu] + \zeta \rangle \\
&= \frac{n}{g(n)} \left\langle \sum_k X_K, \mathbb{E}[\mu] + \zeta \right\rangle \\
&= \frac{n}{g(n)} \langle n\mathbb{E}[\mu] + n\zeta, \mathbb{E}[\mu] + \zeta \rangle \\
&= \frac{n^2}{g(n)} \|\mathbb{E}[\mu] + \zeta\|_2^2 .
\end{aligned}$$

Thus, with an appropriate choice of ϵ,

$$\frac{g(n)}{2n^2 \|\mathbb{E}[\mu]\|_2^2} \leq \frac{1}{\sum_k w_k} \leq \frac{2g(n)}{n^2 \|\mathbb{E}[\mu]\|_2^2},$$

with probability at least $1 - o(1)$. And in particular, if $g(n)$ is $o(n)$, then $\frac{n}{\sum_k w_k} \to 0$. Thus in order to understand the asymptotic behavior of $n \left\| W^{-\frac{1}{2}} T W^{-\frac{1}{2}} \right\|_\infty$ it suffices to understand the asymptotic behavior of $\frac{np_{ij}}{w_i w_j}$. Again noting that $\sum_k X_k = n\mathbb{E}[\mu] + n\zeta$ with high probability, we have that

$$\frac{np_{ij}}{w_i w_j} = \frac{g(n)}{n} \frac{\langle X_i, X_J \rangle}{\langle X_i, \mathbb{E}[\mu] + \zeta \rangle \langle X_j, \mathbb{E}[\mu] + \zeta \rangle}.$$

But then if $\min_i \langle X_i, \mathbb{E}[\mu] \rangle > 0$ with high probability, then $n \left\| W^{-\frac{1}{2}} T W^{-\frac{1}{2}} \right\|_\infty \to 0$.

At this point it is worth noting that if μ satisfies the strong inner product condition, then there is some $\delta > 0$ such that $\mathbb{P}\left(\left\langle \frac{X_i}{\|X_i\|}, \mathbb{E}[\mu] \right\rangle < \delta \right) = 0$, which we will prove in Lemma 4.3. Thus $n \left\| W^{-\frac{1}{2}} T W^{-\frac{1}{2}} \right\|_\infty$ is $o(1)$. However, counterbalancing this, we observe that if $\mathbb{P}(\langle X_i, \mathbb{E}[\mu] \rangle < \delta) = 0$, then there are constants c and c' so that

$$\frac{cg(n)w_{\min}^2}{n} \leq nw_{\min}^2 \left(\frac{1}{\sum_k w_k} + \left\| W^{-\frac{1}{2}} T W^{\frac{-1}{2}} \right\|_\infty \right) \leq \frac{c'g(n)w_{\min}^2}{n}.$$

In other words, if μ satisfies the strong inner product condition $g(n)$ needs to be chosen so that $\frac{n}{g(n)} \geq C_\mu \log^6(n)$ where C_μ is some (potentially large) constant depending on μ.

Thus if $\frac{n}{g(n)} \geq C_\mu \log^6(n)$ and μ satisfies the strong inner product condition, then the second eigenvalue for the normalized adjacency matrix is at most $(1 + o(1)) \|T\| + o(1)$ for $\mathcal{G}_g^{\langle \cdot, \cdot \rangle}(\mu, n)$. However, this leaves open the question of the asymptotic behavior of $\|T\|$. By inspection, it is clear that if μ is supported on only one dimension, then $\|T\| = 0$ and it would be natural to assume that if μ has "low" covariance, then $\|T\|$ would also be small, however we have not been able to prove a result in that direction.

4.1.2 Conductance of $\mathcal{G}_g^{\langle \cdot, \cdot \rangle}(\mu, n)$

In considering the conductance of $\mathcal{G}_g^{\langle \cdot, \cdot \rangle}(\mu, n)$ we are motivated by the work of Flaxman, Frieze, and Vera [28, 29], showing that the geometric preferential attachment model has conductance approaching zero, and further, the conductance is driven towards zero mainly by the cuts induced by the geodesics of the sphere. That is, the worse "bad cuts" in the geometric preferential attachment model are entirely due to the underlying geometry of the model. In this section, we will follow their approach and explore the conductance of $\mathcal{G}_g^{\langle \cdot, \cdot \rangle}(\mu, n)$ via the cuts induced by μ-measurable subsets of $\mathbb{R}^d$. With this end in mind, we define the conductance of a partition $(R, \overline{R})$ as the ratio of the number of edges crossing the cut and the size of the smaller half of the cut. More formally, the regional conductance of a partition is $\frac{C(R,\overline{R})}{\min\left\{\text{Vol}(R),\text{Vol}\left(\overline{R}\right)\right\}}$ where we abuse notation and refer to the set of vertices which lie in a region R by the region itself. In a similar manner, we denote by $\text{Vol}_R(R)$ the volume of the graph induced by the vertices in R. We will also denote by μ_R the probability measure induced by μ on the region R.

Now in order to simplify the notion of the conductance of a region, in the following lemma we characterize $\min\left\{\text{Vol}(R), \text{Vol}\left(\overline{R}\right)\right\}$ in terms of $\mu(R)$, $\mu\left(\overline{R}\right)$, $\mathbb{E}\left[\mu_R\right]$, and $\mathbb{E}\left[\mu_{\overline{R}}\right]$. Before proceeding we note that since $\mu\left(\overline{R}\right) = 1 - \mu(R)$ and $\mathbb{E}\left[\mu_{\overline{R}}\right] = \frac{\mathbb{E}[\mu]+\mu(R)\mathbb{E}[\mu_R]}{1-\mu(R)}$, we may view this characterization strictly in terms of R and μ.

Lemma 4.2. *If* $\mu(R)^2 \left\|\mathbb{E}\left[\mu_R\right]\right\|_2^2 < \mu\left(\overline{R}\right)^2 \left\|\mathbb{E}\left[\mu_{\overline{R}}\right]\right\|_2^2$ *and* $g(n)$ *is* $o(n)$*, then for any* μ*-measurable region* R*,* $\min\left\{\text{Vol}(R), \text{Vol}\left(\overline{R}\right)\right\} = \text{Vol}(R)$ *in* $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ *with probability at least* $1 - o(1)$*.*

Proof. We observe first that $\text{Vol}(R) = C\left(R, \overline{R}\right) + \text{Vol}_R(R)$, and hence it suffices to compare $\text{Vol}_R(R)$ and $\text{Vol}_{\overline{R}}\left(\overline{R}\right)$. Now observe that by Chernoff Bound (C3) for small $\alpha > 0$

$$\mathbb{P}(||R| - \mu(R)\, n| > \alpha\mu(R)\, n) < 2e^{-\frac{\alpha^2\mu(R)n}{3}}.$$

Similarly, if p_R is the mean inner product between vertices in R, then by Chernoff Bound (C3) for small β,

$$\mathbb{P}\left(\left|\text{Vol}_R(R) - \frac{p_R}{g(n)}\binom{|R|}{2}\right| > \beta\frac{p_R}{g(n)}\binom{|R|}{2}\right) \leq e^{-\frac{p_R\binom{|R|}{2}\beta^2}{3g(n)}}.$$

Let $X_1, X_2, \ldots, X_{|R|}$ be the vectors assigned to vertices in R and let $S = \frac{1}{|R|}\sum_i X_i$. Now $p_R = \frac{1}{\binom{|R|}{2}}\sum_{i<j}\langle X_i, X_j\rangle$. Thus we have that

$$\|S\|_2^2 = \frac{1}{|R|^2}\sum_{i,j}\langle X_i, X_j\rangle \geq \frac{2\binom{|R|}{2}p_R}{|R|^2} = \|S\|_2^2 - \frac{\sum_i \|X_i\|_2^2}{|R|^2} \geq \|S\|_2^2 - \frac{1}{|R|}.$$

Thus as $|R| \to \infty$, $\|S\|_2^2 \to p_R$. But now, combining these results with Lemma 4.1, we have that for $\epsilon > 0$,

$$\text{Vol}_R(R) \leq \frac{(1-\beta)\left(\left\|\mathbb{E}\left[\mu_R\right]\right\|_2^2 - \epsilon\right)}{g(n)}\binom{(1-\alpha)\mu(R)\, n}{2},$$

with probability at least $1 - o(1)$. Similarly, with probability at least $1 - o(1)$, we have that

$$\mathrm{Vol}_{\overline{R}}(\overline{R}) \geq \frac{(1+\beta)\left(\left\|\mathbb{E}\left[\mu_{\overline{R}}\right]\right\|_2^2 + \epsilon\right)}{g(n)} \binom{(1+\alpha)\mu(\overline{R})\, n}{2}.$$

But since $\mu(R)^2 \left\|\mathbb{E}\left[\mu_R\right]\right\|_2^2 < \mu(\overline{R})^2 \left\|\mathbb{E}\left[\mu_{\overline{R}}\right]\right\|_2^2$, there is a choice of α, β, and ϵ, so that $\mathrm{Vol}_R(R) < \mathrm{Vol}_{\overline{R}}(\overline{R})$ with probability at least $1 - o(1)$. □

Thus we may consider without loss of generality only μ-measurable regions R for which $\mu(R)^2 \left\|\mathbb{E}\left[\mu_R\right]\right\|_2^2 \leq \mu(\overline{R})^2 \left\|\mathbb{E}\left[\mu_{\overline{R}}\right]\right\|_2^2$. With this in mind, we have the following theorem.

Theorem 4.3. *Let R be a fixed μ-measurable subset of $\mathbb{R}^d$ such that $\mu(R)^2 \left\|\mathbb{E}\left[\mu_R\right]\right\|_2^2 < \mu(\overline{R})^2 \left\|\mathbb{E}\left[\mu_{\overline{R}}\right]\right\|_2^2$. Then, with probability at least $1 - o(1)$,*

$$\frac{C(R, \overline{R})}{\mathrm{Vol}(R)} \geq \frac{\mu(\overline{R}) \left\langle \mathbb{E}\left[\mu_R\right], 2\mathbb{E}\left[\mu_{\overline{R}}\right] \right\rangle}{\left\langle \mathbb{E}\left[\mu_R\right], \mathbb{E}\left[\mu\right] \right\rangle}.$$

Proof. Instead of proving the asymptotic lower bound directly we will show instead that $\frac{g(n)\,\mathrm{Vol}_R(R)}{\binom{\mu(R)n}{2}}$ is not too much larger than 1 and that $\frac{2g(n)C(R,\overline{R})}{\mu(R)\mu(\overline{R})\left\langle \mathbb{E}[\mu_R], \mathbb{E}[\mu_{\overline{R}}]\right\rangle}$ is not too much smaller than 2. In particular

$$\frac{g(n)\,\mathrm{Vol}_R(R)}{\binom{\mu(R)n}{2} \left\|\mathbb{E}\left[\mu_R\right]\right\|_2^2} \leq \frac{2g(n)C(R, \overline{R})}{\mu(R)\,\mu(\overline{R}) \left\langle \mathbb{E}\left[\mu_R\right], \mathbb{E}\left[\mu_{\overline{R}}\right] \right\rangle}$$

and thus

$$\frac{2\mu(R)\,\mu(\overline{R}) \left\langle \mathbb{E}\left[\mu_R\right], \mathbb{E}\left[\mu_{\overline{R}}\right] \right\rangle}{\binom{\mu(R)n}{2} \left\|\mathbb{E}\left[\mu_R\right]\right\|_2^2} \leq \frac{C(R, \overline{R})}{\mathrm{Vol}_R(R)}.$$

Let $\alpha, \beta, \epsilon > 0$ be small constants, then observe that by the same argument as in Lemma 4.2, $\mathrm{Vol}_R(R) \geq (1+\beta)\frac{(\|\mathbb{E}[\mu_R]\|_2^2 + \epsilon)}{g(n)} \binom{(1+\alpha)\mu(R)n}{2}$ with probability at least $1 - o(1)$. Thus for sufficiently small α, β, and ϵ, we have that

$$\mathrm{Vol}_R(R) \leq \frac{3}{2} \left\|\mathbb{E}\left[\mu_R\right]\right\|_2^2 \binom{\mu(R)\, n}{2}. \tag{4.7}$$

Now we note that by Lemma 4.1, with probability at least $1 - o(1)$ the mean inner product between vectors in R and $\overline{R}$ is at least $\left\langle \mathbb{E}\left[\mu_R\right], \mathbb{E}\left[\mu_{\overline{R}}\right] \right\rangle - \epsilon(\left\|\mathbb{E}\left[\mu_R\right]\right\|_2 + \left\|\mathbb{E}\left[\mu_{\overline{R}}\right]\right\|_2) - \epsilon^2$. Further, by Chernoff Bound (C3) and similarly to Lemma 4.2 with probability at $1 - o(1)$, $|R| > (1-\alpha)\mu(R)\, n$ and $\left|\overline{R}\right| > (1-\alpha)\mu(\overline{R})\, n$. Combining these results and applying Chernoff Bound (C3) with parameter β to the occurrence of edges crossing the cut, we have that with probability at least $1 - o(1)$,

$$C(R, \overline{R}) \geq (1-\beta)\frac{\langle \mathbb{E}[\mu_R], \mathbb{E}[\mu_{\overline{R}}]\rangle - \epsilon(\|\mathbb{E}[\mu_R]\|_2 + \|\mathbb{E}[\mu_{\overline{R}}]\|_2) - \epsilon^2}{g(n)}(1-\alpha)^2\mu(R)\,\mu(\overline{R})\,n^2.$$

Thus for sufficiently small α, β, and ϵ,

$$2C(R, \overline{R}) \geq \frac{3}{2}\mu(R)\,\mu(\overline{R})\,n^2\frac{\langle \mathbb{E}[\mu_R], \mathbb{E}[\mu_{\overline{R}}]\rangle}{g(n)}$$

with probability at least $1 - o(1)$. Combining Equation (4.7) and Equation (4.1.2), the result then follows. □

Now in order for the conductance to be constant, there must be some lower bound on $\frac{\mu(\overline{R})\langle \mathbb{E}[\mu_R], 2\mathbb{E}[\mu_{\overline{R}}]\rangle}{\langle \mathbb{E}[\mu_R], \mathbb{E}[\mu]\rangle}$, or alternatively an upper bound on $\frac{\langle \mathbb{E}[\mu_R], \mathbb{E}[\mu]\rangle}{\mu(\overline{R})\langle \mathbb{E}[\mu_R], \mathbb{E}[\mu_{\overline{R}}]\rangle} = 1 + \frac{\mu(R)\|\mathbb{E}[\mu_R]\|_2^2}{\mu(\overline{R})\langle \mathbb{E}[\mu_R], \mathbb{E}[\mu_{\overline{R}}]\rangle}$. To show this, we prove the following lemma about the geometry of μ.

Lemma 4.3. *For any probability distribution μ satisfying the strong inner product condition,*

$$\sup_{(R,\overline{R})} \frac{\mu(R)\,\|\mathbb{E}[\mu_R]\|_2^2}{\mu(\overline{R})\,\langle \mathbb{E}[\mu_R], \mathbb{E}[\mu_{\overline{R}}]\rangle} < \infty,$$

where the supremum is taken over all μ-measurable partitions $(R, \overline{R})$ where

$$\mu(R)^2\,\|\mathbb{E}[\mu_R]\|_2^2 \leq \mu(\overline{R})\,\|\mathbb{E}[\mu_{\overline{R}}]\|_2^2.$$

Proof. We first observe that if θ is the angle between $\mathbb{E}[\mu_R]$ and $\mathbb{E}[\mu_{\overline{R}}]$, then

$$\frac{\mu(R)\,\|\mathbb{E}[\mu_R]\|_2^2}{\mu(\overline{R})\,\langle \mathbb{E}[\mu_R], \mathbb{E}[\mu_{\overline{R}}]\rangle} = \frac{\mu(R)\,\|\mathbb{E}[\mu_R]\|_2}{\mu(\overline{R})\,\|\mathbb{E}[\mu_{\overline{R}}]\|_2\cos(\theta)} \leq \frac{\mu(\overline{R})\,\|\mathbb{E}[\mu_{\overline{R}}]\|_2}{\mu(R)\,\|\mathbb{E}[\mu_R]\|_2\cos(\theta)}.$$

Thus it suffices to show that there is no sequence of μ-measurable sets R_i such that both $\mu(\overline{R_i})\,\left\|\mathbb{E}\left[\mu_{\overline{R_i}}\right]\right\|_2\cos(\theta_i)$ and $\mu(R_i)\,\|\mathbb{E}[\mu_{R_i}]\|_2\cos(\theta_i)$ approach 0, where $\cos(\theta_i)$ is the angle between $\mathbb{E}[\mu_{R_i}]$ and $\mathbb{E}\left[\mu_{\overline{R_i}}\right]$.

Suppose $\{R_i\}$ is such a sequence. By passing to convergent subsequences we may assume without loss of generality that $\mu(R_i)$, $\|\mathbb{E}[\mu_{R_i}]\|_2$, $\left\|\mathbb{E}\left[\mu_{\overline{R_i}}\right]\right\|_2$, and $\cos(\theta_i)$ all converge monotonically. It is clear that if $\cos(\theta_i)$ does not converge to zero, then either $\mu(R_i)$ and $\left\|\mathbb{E}\left[\mu_{\overline{R_i}}\right]\right\|_2$ converge to zero, or $\mu(\overline{R_i})$ and $\|\mathbb{E}[\mu_{R_i}]\|_2$ both converge to zero. However, since

$\mathbb{E}[\mu] = \mu(R_i)\,\mathbb{E}[\mu_{R_i}] + \mu(\overline{R_i})\,\mathbb{E}\left[\mu_{\overline{R_i}}\right]$ it is clear that neither of these cases hold. Thus, we need only concern ourselves with the possibility that $\cos(\theta_i) \to 0$.

To that end we consider the following auxiliary sequences. Define $S_i = \frac{\mathbb{E}[\mu_{R_i}]}{\|\mathbb{E}[\mu_{R_i}]\|_2}$ and $S_i' = \frac{\mathbb{E}\left[\mu_{\overline{R_i}}\right]}{\left\|\mathbb{E}\left[\mu_{\overline{R_i}}\right]\right\|_2}$. Thus by passing to subsequences, we may assume that there exist some S^* and S'^* such that $S_i \to S^*$ and $S_i' \to S'^*$. Furthermore, by the application of an orthogonal matrix Q we may assume that $QS^* = e_1$ and that there are some α and β so that $QS'^* = \alpha e_1 + \beta e_2$. Now by assumption, $\langle S_i, S_i' \rangle$ approaches 0 and thus $\alpha = 0$ and $\beta = \pm 1$. Without loss of generality, we suppose $\beta = 1$. Further, since Q is orthonormal, we may consider without loss of generality the alternative measure $Q\mu$, and thus $S^* = e_1$ and $S'^* = e_2$.

Let $R' = \left\{x \in \mathbb{R}^d \mid \langle x, e_2 \rangle < 0\right\}$. Suppose now that $\mu(R') > 0$. But then, since $S_i' \to e_2$, for sufficiently large i, $\langle \mathbb{E}[\mu_{R'}], S_i \rangle < 0$, which contradicts the inner product condition since this is a rescaling of $\left\langle \mathbb{E}[\mu_{R'}], \mathbb{E}\left[\mu_{\overline{R_i}}\right] \right\rangle$. Thus $\mu(R') = 0$. In a similar manner, we have that $\mu\left(\left\{x \in \mathbb{R}^d \mid \langle x, e_1 \rangle < 0\right\}\right) = 0$.

Now let

$$\begin{aligned} p &= \lim_{i \to \infty} \mu(R_i)\,, \\ R^* &= \left\{x \in \mathbb{R}^d \mid \langle x, e_2 \rangle = 0\right\}, \qquad \text{and} \\ R'^* &= \left\{x \in \mathbb{R}^d \mid \langle x, e_1 \rangle = 0\right\}. \end{aligned}$$

Suppose $p > \mu(R^*)$. Then for sufficiently large i, since $\langle \mathbb{E}[\mu_{R_i}], e_2 \rangle \to 0$ there is μ-measurable set $R_i' \subseteq R_i$ such that $\left\langle \mathbb{E}\left[\mu_{R_i'}\right], e_2 \right\rangle < 0$ and $\mu(R_i') > 0$. But this contradicts that $\mu(R') = 0$. Thus we may assume that $p \leq \mu(R^*)$. In a similar manner, we may assume that $(1 - p) \leq \mu(R'^*)$. Furthermore, since $\langle \mathbb{E}[\mu_{R^* \cap R'^*}], \mathbb{E}[\mu] \rangle = 0$, we may assume that $\mu(R^* \cap R'^*) = 0$. Thus $\mu(R^*) = p$ and $\mu(R'^*) = 1 - p$. But then since $\mu\left(R_i \cap \left\{x \in \mathbb{R}^d \mid \langle x, e_2 \rangle > 0\right\}\right) \to 0$, in a probabilistic sense R_i approaches R^*. Similarly, $\overline{R_i}$ approaches R'^*. But then $\mathbb{E}[\mu_{R^*}] = \lambda e_1$ and $\mathbb{E}[\mu_{R'^*}] = \lambda' e_2$. Thus $\langle \mathbb{E}[\mu_{R^*}], \mathbb{E}[\mu_{R'^*}] \rangle = 0$, and so if $0 < p < 1$ this contradicts the strong inner product condition.

Thus we may assume without loss of generality that $p = 1$. Then $\frac{\mathbb{E}[\mu]}{\|\mathbb{E}[\mu]\|_2} = \frac{\mathbb{E}[\mu_{R^*}]}{\|\mathbb{E}[\mu_{R^*}]\|_2} = e_1$. In addition, since $S_i' \to e_2$, we have that for sufficiently large i, $\langle S_i', e_2 \rangle > 0$. But then since

$\langle \mathbb{E}[\mu], e_2 \rangle = 0$, we have $\langle \mathbb{E}[\mu_{R_i}], e_2 \rangle < 0$, a contradiction.

Thus there is $c > 0$ such that for any sequence of $\{R_i\}$, $\left\langle \frac{\mathbb{E}[\mu_{R_i}]}{\|\mathbb{E}[\mu_{R_i}]\|_2}, \frac{\mathbb{E}\left[\mu_{\overline{R_i}}\right]}{\left\|\mathbb{E}\left[\mu_{\overline{R_i}}\right]\right\|_2} \right\rangle \geq c$, and in particular there is a $c' > 0$ such that at least one of $\mu(\overline{R_i}) \left\|\mathbb{E}\left[\mu_{\overline{R_i}}\right]\right\|_2 \cos(\theta_i)$ and $\mu(R_i) \|\mathbb{E}[\mu_{R_i}]\|_2 \cos(\theta_i)$ is larger than c'. □

Combining Lemma 4.3 with Theorem 4.3, there is some constant, depending only on μ, such that for any μ-measurable region the conductance of the cut induced by the region is at least the constant. We note that, since there are exponentially many relevant regions for any n, this methodology does not extend to show that the conductance in general is constant. However, this does further the intuition that if there are cuts with small conductance, one side of the cut consists of a small ($o(n)$) portion of the total graph. This is in contrast to the geometric preferential attachment model of Flaxman, Frieze, and Vera [28, 29] and further aligns with empirical observations of community structure in complex networks [56, 43, 44].

4.2 Assortativity

We consider here the assortativity of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ as a function of μ. The assortativity of a graph is designed to be an empirical means of quantifying the nature of the connections in the graph, in particular it is designed as a means of estimating the "second-order" degree distribution. That is, if the assortativity is high then the intuition is that the vertices of high degree tend to have other vertices of high degree as neighbors. On the other hand, if the assortativity is low then vertices of high degree tend to have mostly vertices of low degree in their neighborhood. There are a few competing means of measuring assortativity, or an assortativity-like statistic, that appear in the physics literature. We will briefly review these measures and argue for one of them as superior from a graph theoretic point of view.

As a basis for terminology, fix some graph G and let p_i be the fraction of the vertices that have degree i, and let the sequence of degrees be $d_1 < d_2 < \cdots < d_n$. Further, define d_{jk} to be the fraction of edges in G that go between a vertex of degree j and a vertex of degree k. In addition, we define $q_k = \frac{(k+1)p_{k+1}}{\sum_j jp_j}$ and $\sigma_q^2 = \sum_k k^2 q_k - (\sum_k kq_k)^2$. With these definitions Newman defines the assortativity as

$$\frac{1}{\sigma_q^2}\sum_{j,k} jk(d_{jk} - q_j q_k).$$

Interpreting q_k as the randomized "remaining degree" on leaving a given vertex, Newman claims that σ_q^2 is maximal value of the degree-degree correlation function on the given degree sequence [49]. Thus this definition of assortativity ranges from -1 to 1, with a value of 0 representing neutral assortativity, positive values representing an assortative network and negative values representing a disassortative network.

On the other hand, Alderson and Li [3] define the simpler measure

$$\mathcal{D}(G) = \sum_{u,v \in E(G)} \deg(u)\deg(v) \tag{4.8}$$

They also note that Newman's assortativity can be derived from (4.8) by a linear transformation depending on the degree sequence. The measure of Alderson and Li has the advantage of being a natural graph theoretic measure, but lacks the advantage of a natural critical value that serves as the transition from disassortative to assortative networks. However, since the Erdős-Rényí is clearly neutrally assortative, by calculating the assortativity for a specified edge density we can find an estimate of where the transition between positive and negative assortativity lies.

One further advantage of the Newman definition of assortativity is that it is normalized between -1 and 1, and so makes comparison between various graphs more natural. However, the normalization used fixes the degree distribution of the graph, which severely limits the broad applicability of assortativity as a means of comparison between different networks, let alone different classes of network. Newman calculates the assortativity of several types of sociological networks and biological networks in [49], but other than the sign, these number are of little use. For instance, the value of assortativity for the physics, biology, and mathematics co-authorership networks are 0.363, 0.127, and 0.120, respectively. It would be tempting to conclude from these values, that in some significant way biologists and mathematicians collaborate in a similar but different way than physicists collaborate. But since the normalization is preformed with respect to the degree distribution the proximity of these values mean nothing. Furthermore, normalizing with respect to the degree distribution precludes the possibility of analyzing the change in assortativity as a network evolves, since the measures have little relationship to each other if the degree distribution changes, or even worse, the number of nodes changes. For this reason, we choose not to normalize assortativity, using instead a structural version of the definition of Alderson and Li. In particular, we determine asymptotic lower and upper bounds on $\mathcal{D}(G)$ in terms of the number of edges, and the asymptotic behavior of Erdős-Rényí in terms of edge density.

Lemma 4.4. *For a graph G with m edges, $m \leq \mathcal{D}(G) \leq 2m^2$, and further, these bounds are asymptotically tight.*

Proof. First note that $\mathcal{D}(G) \geq \sum_{\{u,v\}\in E(G)} 1 = m$. But it is clear that there exists a graph on m edges and at least $2m$ vertices such that $\mathcal{D}(G) = m$ and so the lower bound is tight. In a similar manner

$$\begin{aligned}\mathcal{D}(G) &= \sum_{\{u,v\}\in E(G)} \deg(u)\deg(v) \\ &\leq \sum_{u,v\in V(G)} \deg(u)\deg(v) \\ &= \frac{1}{2}\left(\left(\sum_u \deg(u)\right)^2 - \sum_v \deg(v)^2\right) \\ &= 2m^2 - \frac{1}{2}\sum_v \deg(v)^2 \\ &\leq 2m^2.\end{aligned}$$

Now let k and t be naturals such that $0 \leq t \leq k$ and $\binom{k}{2} + t = m$. Let H be the graph formed on $[k+1]$ by letting the graph induced by $[k]$ be complete and letting the degree of the vertex $k+1$ be t, then H has m edges. Furthermore

$$\mathcal{D}(H) = \binom{k-t}{2}(k-1)^2 + (k-t)tk(k-1) + \binom{t}{2}k^2 + t^2k.$$

One can quickly observe that for fixed k the minimum either occurs at $t = 0$ or $t = k$, and a cursory inspection yields that the minimum occurs at $t = 0$. Thus

$$\mathcal{D}(H) \geq \binom{k}{2}(k-1)^2 = 2\binom{k}{2}^2 - \binom{k}{2}(k-1) = 2(m-t)^2 - (m-t)(k-1).$$

However, since $t \leq k \leq 2\sqrt{m}$, this behaves asymptotically as $2m^2$. $\square$

We now move to a structural interpretation of $\mathcal{D}(G)$ that will facilitate the analysis of $\mathcal{D}(G)$ for random graphs.

Lemma 4.5. *$\mathcal{D}(G)$ for an undirected graph G is equal to half the number of directed walks of length three.*

Proof. Let A be the adjacency matrix for G, then we note that

$$\mathcal{D}(G) = \sum_{\{i,j\}\in E(G)} \deg(i)\deg(j) = \frac{1}{2}\sum_{i,j}\deg(i)\, a_{ij}\deg(j) = \sum_{u,i,j,v} a_{ui}a_{ij}a_{jv} = \frac{1}{2}\mathbb{1}^T A^3 \mathbb{1}.$$

This is well known to be half the number of directed walks of length three. $\square$

This leads immediately to the following structural definition of assortativity that is significantly easier to apply to a wide class of random graph models.

Corollary 4.1. *If E is the number of edges of a graph, P_2 is the number of paths of length 2, T the number of triangles, and P_3 the number of paths of length 3, then $\mathcal{D}(G) = E + 2P_2 + 3T + P_3$.*

Proof. Note that a directed walk of length 3 traverses at most 3 edges and so the subgraphs induced by these walk are precisely the edges, the paths of length 2, the paths of length 3, and the triangles. Now for the single edge, it is clear there are only 2 directed paths, one starting at each endpoint. For the triangles, there are 3 potential starting points and 2 directions, so there are 6 directed walks. For the path of length 2, there are 4 directed walks; one starting at each endpoint, and two starting at the center endpoint. Finally for the path of length 3, there are 2 directed walks, one starting at each endpoint. Thus the total number of directed walks is, $2E+4P_2+6T+2P_3$, and so $\mathcal{D}(G) = E+2P_2+3T+P_3$. $\square$

Using this result we can quickly characterize the assortativity the Erdős-Rényí graphs which are clearly neutrally assortative. Thus we have a point of comparison for assortativity or disassortativity while maintaining a relatively easy formula.

Corollary 4.2.

$$\mathbb{E}\left[\mathcal{D}\left(\mathcal{G}(n,p)\right)\right] = \binom{n}{2}\left(p + (2n-4)p^2 + (n-2)(2n-2)p^3\right).$$

$$\mathbb{E}\left[\mathcal{D}\left(\mathcal{G}(n,m)\right)\right] = m + \frac{4m(m-1)}{n+1}\left(1 + \frac{(n-2)(m-2)}{n^2-n-4}\right).$$

Proof. Let E, P_2, P_3, and T be defined as in Corollary 4.1. Because of the independence of the edges in $\mathcal{G}(n,p)$, it is immediate that

$$\mathbb{E}\left[E\right] = \binom{n}{2}p \qquad \mathbb{E}\left[T\right] = \binom{n}{3}p^3$$
$$\mathbb{E}\left[P_2\right] = \binom{n}{2}(n-2)p^2 \qquad \mathbb{E}\left[P_3\right] = \binom{n}{2}\binom{n-2}{2}4p^3.$$

Thus

$$\begin{aligned}\mathbb{E}\left[\mathcal{D}\left(\mathcal{G}(n,p)\right)\right] &= \mathbb{E}\left[E\right] + 2\mathbb{E}\left[P_2\right] + \mathbb{E}\left[P_3\right] + 3\mathbb{E}\left[T\right] \\ &= \binom{n}{2}\left(p + 2(n-2)p^2 + \binom{n-2}{2}4p^3 + (n-2)p^3\right) \\ &= \binom{n}{2}\left(p + (2n-4)p^2 + (n-2)(2n-5)p^3\right).\end{aligned}$$

In a similar manner, for $\mathcal{G}(n,m)$,

$$\mathbb{E}\left[E\right] = m \qquad \mathbb{E}\left[T\right] = \frac{4m(m-1)(m-2)}{3(n^2-n-4)(n+1)}$$
$$\mathbb{E}\left[P_2\right] = \frac{2m(m-1)}{n+1} \qquad \mathbb{E}\left[P_3\right] = \frac{4m(m-1)(m-2)(n-3)}{(n+1)(n^2-n-4)}.$$

Thus $\mathbb{E}\left[\mathcal{D}\left(\mathcal{G}(n,m)\right)\right] = m + \frac{4m(m-1)}{n+1}\left(1 + \frac{(n-2)(m-2)}{n^2-n-4}\right)$. □

This leads us to the main result of this section, that $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu,n)$ exhibits positive assortativity and that as $n \to \infty$ the assortativity of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu,n)$ tends towards neutral.

Theorem 4.4. *The assortativity of $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu,n)$ is positive, and in particular*

$$\frac{\mathbb{E}\left[\mathcal{D}\left(\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu,n)\right)\right] - \mathbb{E}\left[\mathcal{D}\left(\mathcal{G}\left(n, \frac{\|\mathbb{E}[\mu]\|^2}{g(n)}\right)\right)\right]}{2\mathbb{E}\left[\left|E(\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu,n))\right|\right]^2}$$

is $\Theta\left(\frac{1}{n} + \frac{1}{g(n)} + \frac{g(n)}{n^2}\right)$.

Proof. For notational convenience let $p = \frac{\|\mathbb{E}[\mu]\|^2}{g(n)}$ and let $p' = \|\mathbb{E}[\mu]\|^2$. Let E be the expected number of edges, T be the expected number of triangles, and P_2 and P_3 be the expected number of paths of length 2 and length 3, respectively, for $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$. It is clear that $E = p\binom{n}{2}$, so we now proceed to determine the values for T, P_3 and P_2.

$$
\begin{aligned}
T &= \binom{n}{3} \iiint \frac{\langle X_u, X_v\rangle \langle X_v, X_w\rangle \langle X_w, X_u\rangle}{g(n)^3} d\mu(X_u) d\mu(X_v) d\mu(X_w) \\
&= \binom{n}{3} \int \frac{X_u^T \left(\operatorname{cov}(\mu) + \mathbb{E}[\mu]\mathbb{E}[\mu]^T\right)^2 X_u}{g(n)^3} d\mu(X_u) \\
&= \binom{n}{3} \int \frac{X_u^T \operatorname{cov}(\mu)^2 X_u + X_u^T \operatorname{cov}(\mu)\mathbb{E}[\mu]\mathbb{E}[mu]^T X_u + X_u^T \mathbb{E}[\mu]\mathbb{E}[\mu]^T \operatorname{cov}(\mu) X_u}{g(n)^3} d\mu(X_u) \\
&\quad + \binom{n}{3} \int \frac{p \langle X_u, \mathbb{E}[\mu]\rangle^2}{g(n)^3} d\mu(X_u) \\
&= \binom{n}{3} \int \frac{X_u^T \operatorname{cov}(\mu)^2 X_u + X_u^T \operatorname{cov}(\mu)\mathbb{E}[\mu]\mathbb{E}[mu]^T X_u + X_u^T \mathbb{E}[\mu]\mathbb{E}[\mu]^T \operatorname{cov}(\mu) X_u}{g(n)^3} d\mu(X_u) \\
&\quad + \binom{n}{3} \frac{p\mathbb{E}[\mu]^T (\operatorname{cov}(\mu) + \mathbb{E}[\mu]\mathbb{E}[\mu]^T)\mathbb{E}[\mu]}{g(n)^3} \\
&= \binom{n}{3} \int \frac{X_u^T \operatorname{cov}(\mu)^2 X_u + X_u^T \operatorname{cov}(\mu)\mathbb{E}[\mu]\mathbb{E}[mu]^T X_u + X_u^T \mathbb{E}[\mu]\mathbb{E}[\mu]^T \operatorname{cov}(\mu) X_u}{g(n)^3} d\mu(X_u) \\
&\quad + \binom{n}{3} \frac{p'\mathbb{E}[\mu]^T (\operatorname{cov}(\mu) + \mathbb{E}[\mu]\mathbb{E}[\mu]^T)\mathbb{E}[\mu]}{g(n)^3} \\
&= \binom{n}{3} \int \frac{X_u^T \operatorname{cov}(\mu)^2 X_u + X_u^T \operatorname{cov}(\mu)\mathbb{E}[\mu]\mathbb{E}[mu]^T X_u + X_u^T \mathbb{E}[\mu]\mathbb{E}[\mu]^T \operatorname{cov}(\mu) X_u}{g(n)^3} d\mu(X_u) \\
&\quad + \binom{n}{3} \left(p \frac{\mathbb{E}[\mu]^T \operatorname{cov}(\mu)\mathbb{E}[\mu]}{g(n)^2} + p^3 \right).
\end{aligned}
$$

$$
\begin{aligned}
P_3 &= \binom{n}{2}(n-2)(n-3) \iiiint \frac{\langle X_u, X_v\rangle \langle X_v, X_w\rangle \langle X_w, X_z\rangle}{g(n)^3} d\mu(X_u) d\mu(X_v) d\mu(X_w) d\mu(X_z) \\
&= \binom{n}{2}(n-2)(n-3) \int \frac{\langle \mathbb{E}[\mu], X_v\rangle \langle X_v, X_w\rangle \langle X_w, \mathbb{E}[\mu]\rangle}{g(n)^3} d\mu(X_v) d\mu(X_w) \\
&= \binom{n}{2}(n-2)(n-3) \frac{\mathbb{E}[\mu]^T (\operatorname{cov}(\mu) + \mathbb{E}[\mu]\mathbb{E}[\mu]^T)^2 \mathbb{E}[\mu]}{g(n)^3} \\
&= \binom{n}{2}(n-2)(n-3) \left(\frac{\mathbb{E}[\mu]^T \operatorname{cov}(\mu)^2 \mathbb{E}[\mu] + 2p'\mathbb{E}[\mu]^T \operatorname{cov}(\mu)\mathbb{E}[\mu] + p'^3}{g(n)^3} \right) \\
&= \binom{n}{2}(n-2)(n-3) \left(\frac{\mathbb{E}[\mu]^T \operatorname{cov}(\mu)^2 \mathbb{E}[\mu] + 2p'\mathbb{E}[\mu]^T \operatorname{cov}(\mu)\mathbb{E}[\mu]}{g(n)^3} + p^3 \right).
\end{aligned}
$$

$$
\begin{aligned}
P_2 &= \binom{n}{2}(n-2)\iiint \frac{\langle X_u, X_v\rangle \langle X_v, X_w\rangle}{g(n)^2} d\mu(X_u)d\mu(X_v)d\mu(X_w) \\
&= \binom{n}{2}(n-2)\int \frac{\langle \mathbb{E}[\mu], X_v\rangle \langle X_v, \mathbb{E}[\mu]\rangle}{g(n)^2} d\mu(X_v) \\
&= \binom{n}{2}(n-2)\frac{\mathbb{E}[\mu]^T (\operatorname{cov}(\mu) + \mathbb{E}[\mu]\mathbb{E}[\mu]^T)\mathbb{E}[\mu]}{g(n)^2} \\
&= \binom{n}{2}(n-2)\left(\frac{\mathbb{E}[\mu]^T \operatorname{cov}(\mu)\,\mathbb{E}[\mu]}{g(n)^2} + p^2\right).
\end{aligned}
$$

Now, as in the proof of Theorem 2.2, we may assume that $\operatorname{cov}(\mu)$ is diagonal, and in particular $\operatorname{cov}(\mu_i\mu_j) = 0$ unless $i = j$. From this we observe that

$$
\begin{aligned}
\int X_u^T \operatorname{cov}(\mu)^2 X_u d\mu(X_u) &= \sum_i \operatorname{Var}(\mu_i)^2\, \mathbb{E}\left[\mu_i^2\right] \\
&= \sum_i \operatorname{Var}(\mu_i)^3 + \mathbb{E}[\mu_i]^2 \operatorname{Var}(\mu_i)^2 \\
&= \mathbb{1}^T \operatorname{cov}(\mu)^3\, \mathbb{1} + \mathbb{E}[\mu]^T \operatorname{cov}(\mu)^2\, \mathbb{E}[\mu].
\end{aligned}
$$

In a similar manner,

$$
\begin{aligned}
\int X_u^T \operatorname{cov}(\mu)\,\mathbb{E}[\mu]\,\mathbb{E}[mu]^T X_u d\mu(X_u) &= \mathbb{E}[\mu]^T \operatorname{cov}(\mu)^2\, \mathbb{E}[\mu] + p'\mathbb{E}[\mu]^T \operatorname{cov}(\mu)\, \mathbb{E}[\mu] \\
&= \int X_u^T \mathbb{E}[\mu]\,\mathbb{E}[\mu]^T \operatorname{cov}(\mu)\, X_u d\mu(X_u).
\end{aligned}
$$

Thus

$$
T = \binom{n}{3}\left(\frac{\mathbb{1}^T \operatorname{cov}(\mu)\,\mathbb{1} + 3\mathbb{E}[\mu]^T \operatorname{cov}(\mu)^2\, \mathbb{E}[\mu]}{g(n)^3} + 3p\frac{\mathbb{E}[\mu]^T \operatorname{cov}(\mu)\,\mathbb{E}[\mu]}{g(n)^2} + p^3\right).
$$

Now combining these results with Corollary 4.1, Corollary 4.2, and the observation that $\operatorname{cov}(\mu)$ is positive semi-definite, yields that $\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu, n)$ has positive assortativity. Now we observe that

$$
\begin{aligned}
\frac{E}{2E^2} &= \Omega\left(\frac{g(n)}{n^2}\right) \\
\frac{P_2}{2E^2} &= \Omega\left(\frac{1}{n}\right) \\
\frac{P_3}{2E^2} &= \Omega\left(\frac{1}{g(n)}\right) \\
\frac{T}{2E^2} &= \Omega\left(\frac{1}{g(n)n}\right).
\end{aligned}
$$

Thus $\frac{\mathbb{E}\left[\mathcal{D}\left(\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu,n)\right)\right] - \mathbb{E}\left[\mathcal{D}\left(\mathcal{G}\left(n, \frac{\|\mathbb{E}[\mu]\|^2}{g(n)}\right)\right)\right]}{2\mathbb{E}\left[\left|E(\mathcal{G}_g^{\langle\cdot,\cdot\rangle}(\mu,n))\right|\right]^2}$ is $\Omega\left(\frac{1}{n} + \frac{1}{g(n)} + \frac{g(n)}{n^2}\right)$. □

REFERENCES

[1] "Bob metcalfe eats his words," *Internet Computing Online*, vol. 1, no. 3, 1997.

[2] Achlioptas, D., Clauset, A., Kempe, D., and Moore, C., "On the bias of traceroute sampling: or, power-law degree distributions in regular graphs," in *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, (New York, NY, USA), pp. 694–703, ACM, 2005.

[3] Alderson, D. L. and Li, L., "Diversity of graphs with highly variable connectivity," *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, vol. 75, no. 4, p. 046102, 2007.

[4] Alon, N. and Spencer, J. H., *The probabilistic method.* Wiley-Interscience Series in Discrete Mathematics and Optimization, New York: Wiley-Interscience [John Wiley & Sons], second ed., 2000. With an appendix on the life and work of Paul Erdős.

[5] Azar, Y., Fiat, A., Karlin, A., McSherry, F., and Saia, J., "Spectral analysis of data," in *STOC '01: Proceedings of the thirty-third annual ACM symposium on Theory of computing*, (New York, NY, USA), pp. 619–626, ACM, 2001.

[6] Barabási, A.-L., *Linked: The New Science of Networks.* Perseus Publishing, 2002.

[7] Barabási, A.-L. and Albert, R., "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[8] Ben-Tal, A. and Nemirovski, A., *Lectures on modern convex optimization.* MPS/SIAM Series on Optimization, Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2001. Analysis, algorithms, and engineering applications.

[9] Bollobás, B., "The diameter of random graphs," *Trans. Amer. Math. Soc.*, vol. 267, no. 1, pp. 41–52, 1981.

[10] Bollobás, B., "The evolution of random graphs," *Trans. Amer. Math. Soc.*, vol. 286, no. 1, pp. 257–274, 1984.

[11] Bollobás, B., *Modern graph theory*, vol. 184 of *Graduate Texts in Mathematics.* New York: Springer-Verlag, 1998.

[12] Bollobás, B., *Random graphs*, vol. 73 of *Cambridge Studies in Advanced Mathematics.* Cambridge: Cambridge University Press, second ed., 2001.

[13] Bollobás, B., Janson, S., and Riordan, O., "The phase transition in inhomogeneous random graphs," *Random Structures Algorithms*, vol. 31, no. 1, pp. 3–122, 2007.

[14] Bollobás, B. and Riordan, O., "The diameter of a scale-free random graph," *Combinatorica*, vol. 24, no. 1, pp. 5–34, 2004.

[15] Bollobás, B., Riordan, O., Spencer, J., and Tusnády, G., "The degree sequence of a scale-free random graph process," *Random Structures Algorithms*, vol. 18, no. 3, pp. 279–290, 2001.

[16] Bornholdt, S. and Schuster, H. G., eds., *Handbook of graphs and networks.* Wiley-VCH, Weinheim, 2003. From the genome to the internet.

[17] Caldarelli, G., Capocci, A., De Los Rios, P., and Muñoz, M. A., "Scale-free networks from varying vertex intrinsic fitness," *Phys. Rev. Lett.*, vol. 89, p. 258702, Dec 2002.

[18] Chung, F. and Lu, L., "Connected components in random graphs with given expected degree sequences," *Ann. Comb.*, vol. 6, no. 2, pp. 125–145, 2002.

[19] Chung, F. and Lu, L., "The average distance in a random graph with given expected degrees," *Internet Math.*, vol. 1, no. 1, pp. 91–113, 2003.

[20] Chung, F. and Lu, L., *Complex graphs and networks*, vol. 107 of *CBMS Regional Conference Series in Mathematics.* Published for the Conference Board of the Mathematical Sciences, Washington, DC, 2006.

[21] Chung, F., Lu, L., Dewey, G. T., and Galas, D. J., "Duplication models for biological networks," *Journal of Computational Biology*, vol. 10, no. 5, pp. 677–688, 2003.

[22] Chung, F., Lu, L., and Vu, V., "The spectra of random graphs with given expected degrees," *Internet Math.*, vol. 1, no. 3, pp. 257–275, 2004.

[23] Chung, F. R. K., *Spectral graph theory*, vol. 92 of *CBMS Regional Conference Series in Mathematics.* Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1997.

[24] Durrett, R., *Random graph dynamics.* Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge: Cambridge University Press, 2007.

[25] Estrada, E., "Spectral scaling and good expansion properties in complex networks," *EPL (Europhysics Letters)*, vol. 73, no. 4, pp. 649–655, 2006.

[26] Evans, L. C. and Gariepy, R. F., *Measure theory and fine properties of functions.* Studies in Advanced Mathematics, Boca Raton, FL: CRC Press, 1992.

[27] Faloutsos, M., Faloutsos, P., and Faloutsos, C., "On power-law relationships of the internet topology," in *SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, (New York, NY, USA), pp. 251–262, ACM, 1999.

[28] Flaxman, A. D., Frieze, A. M., and Vera, J., "A geometric preferential attachment model of networks," *Internet Math.*, vol. 3, no. 2, pp. 187–205, 2006.

[29] Flaxman, A. D., Frieze, A. M., and Vera, J., "A geometric preferential attachment model of networks ii," in *Algorithms and Models for the Web-Graph* (Chung, F. and Bonato, A., eds.), vol. 4863 of *Lecture Notes in Computer Science*, pp. 41–55, 2007.

[30] Folland, G. B., *Real analysis.* Pure and Applied Mathematics (New York), New York: John Wiley & Sons Inc., second ed., 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[31] Füredi, Z. and Komlós, J., "The eigenvalues of random symmetric matrices," *Combinatorica*, vol. 1, no. 3, pp. 233–241, 1981.

[32] Gkantsidis, C., Mihail, M., and Saberi, A., "Conductance and congestion in power law graphs," *SIGMETRICS Perform. Eval. Rev.*, vol. 31, no. 1, pp. 148–159, 2003.

[33] Jackson, M. O., *Social and economic networks.* Princeton, NJ: Princeton University Press, 2008.

[34] Kleinberg, J., "The small-world phenomenon: an algorithm perspective," in *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, (New York, NY, USA), pp. 163–170, ACM, 2000.

[35] Kleinberg, J., "Complex networks and decentralized search algorithms," in *International Congress of Mathematicians. Vol. III*, pp. 1019–1044, Eur. Math. Soc., Zürich, 2006.

[36] Kraetzl, M., Nickel, C., and Scheinerman, E. R., "Random dot product graphs: A model for social networks." 2005.

[37] Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A., and Upfal, E., "Stochastic models for the web graph," *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pp. 57–65, 2000.

[38] Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tompkins, A., and Upfal, E., "The web as a graph," in *PODS '00: Proceedings of the nineteenth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, (New York, NY, USA), pp. 1–10, ACM, 2000.

[39] Lakhina, A., Byers, J., Crovella, M., and Xie, P., "Sampling biases in ip topology measurements," *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE*, vol. 1, pp. 332–341 vol.1, March-3 April 2003.

[40] Leskovec, J., Chakrabarti, D., Kleinberg, J., and Faloutsos, C., "Realistic, mathematically tractable graph generation and evolution, using kronecker multiplication," in *European Conference on Principles and Practice of Knowledge Discovery in Database*, 2005.

[41] Leskovec, J. and Faloutsos, C., "Scalable modeling of real graphs using kronecker multiplication," in *ICML '07: Proceedings of the 24th international conference on Machine learning*, (New York, NY, USA), pp. 497–504, ACM, 2007.

[42] Leskovec, J., Kleinberg, J., and Faloutsos, C., "Graph evolution: Densification and shrinking diameters," *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 1, p. 2, 2007.

[43] Leskovec, J., Lang, K. J., Dasgupta, A., and Mahoney, M. W., "Community structure in large networks: Natural cluster sizes and the absence of large well-define clusters," tech. rep., 2008.

[44] LESKOVEC, J., LANG, K. J., DASGUPTA, A., and MAHONEY, M. W., "Statistical properties of community structure in large social and information networks," in *WWW '08: Proceeding of the 17th international conference on World Wide Web*, (New York, NY, USA), pp. 695–704, ACM, 2008.

[45] MAHDAIN, M. and XU, Y., "Stochastic kronecker graphs," in *Algorithms and Models for the Web-Graph* (CHUNG, F. and BONATO, A., eds.), vol. 4863 of *Lecture Notes in Computer Science*, pp. 179–186, 2007.

[46] MIHAIL, M., PAPADIMITRIOU, C., and SABERI, A., "On certain connectivity properties of the internet topology," *J. Comput. System Sci.*, vol. 72, no. 2, pp. 239–251, 2006.

[47] MOLLOY, M. and REED, B., "A critical point for random graphs with a given degree sequence," in *Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science, "Random Graphs '93" (Poznań, 1993)*, vol. 6, pp. 161–179, 1995.

[48] MOLLOY, M. and REED, B., "The size of the giant component of a random graph with a given degree sequence," *Combin. Probab. Comput.*, vol. 7, no. 3, pp. 295–305, 1998.

[49] NEWMAN, M. E. J., "Assortative mixing in networks," *Phys. Rev. Lett.*, vol. 89, p. 208701, Oct 2002.

[50] NEWMAN, M. E. J., "The structure and function of complex networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167–256 (electronic), 2003.

[51] SCHÖLKOPF, B. and SMOLA, A. J., *Learning with Kernels*. MIT Press, 2001.

[52] SHAWE-TAYLOR, J. and CHRISTIANINI, N., *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[53] SINCLAIR, A., *Algorithms for random generation and counting*. Progress in Theoretical Computer Science, Boston, MA: Birkhäuser Boston Inc., 1993. A Markov chain approach.

[54] SÖDERBERG, B., "General formalism for inhomogeneous random graphs," *Phys. Rev. E (3)*, vol. 66, no. 6, pp. 066121, 6, 2002.

[55] VAZIRANI, V. V., *Approximation algorithms*. Berlin: Springer-Verlag, 2001.

[56] WATTS, D. J. and STROGATZ, S. H., "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 1998.

[57] YOUNG, S. J., "The conductance of stochastic kronecker graphs via a general spectral theorem." in preperation.

[58] YOUNG, S. J. and MIHAIL, M., "General random dot product graphs." manuscript, March 2008.

[59] YOUNG, S. J. and SCHEINERMAN, E. R., "Random dot product graph models for social networks," in *Algorithms and Models for the Web-Graph* (CHUNG, F. and BONATO, A., eds.), vol. 4863 of *Lecture Notes in Computer Science*, pp. 138–149, 2007.

[60] Young, S. J. and Scheinerman, E. R., "Directed random dot product graphs," *Internet Math.*, accepted 2008.