CHALLENGES IN RANDOM GRAPH MODELS WITH DEGREE HETEROGENEITY: EXISTENCE, ENUMERATION AND ASYMPTOTICS OF THE SPECTRAL RADIUS

by

David E. Burstein

B.S. Mathematics, University of Maryland, 2011

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

David E. Burstein

It was defended on

June 23rd 2016

and approved by

Jonathan E. Rubin, PhD, Professor, Department of Mathematics

Gregory M. Constantine, PhD, Professor, Department of Mathematics

Brent Doiron, PhD, Associate Professor, Department of Mathematics

Satish Iyengar, PhD, Professor, Department of Statistics

Dissertation Director: Jonathan E. Rubin, PhD, Professor, Department of Mathematics

PREFACE

With great gratitude, I am much relieved to complete this edgy piece of mathematics. And of course, I could not have completed this work without much support from the broader academic community. First off, I would like to thank Jeffrey Wheeler. I have thoroughly enjoyed our mathematically graphic conversations and am greatly appreciative of your guidance and supervision throughout the process of proposing and defending my thesis. It goes without saying that the not so simple path to writing a thesis can at times be long and daunting. I can definitely say thorough that your sense of humor and amiability has made it an enjoyable experience.

I would also like to also give a special thanks to one of my graduate school professors and committee members, Satish Iyengar. I can *almost surely* say that listening and participating in your lectures have tremendously enhanced my graduate school experience. I am incredibly grateful that you agreed to be a part of my thesis comittee, even though you certainly had a substantial number of other commitments while going on your sabbatical. It has been nothing short of an absolute pleasure working with you. I am very much indebted to my thesis committee members, Gregory Constantine and Brent Doiron; I have especially enjoyed the big-picture questions you have asked and feel that the motivation behind this work would not be nearly as enjoyable without your thought-provoking inquiries.

At this juncture, I would like to give thanks to my collaborators, the researchers at the Software Engineering Institute at Carnegie Mellon University and the participants and organizers of the 2014 AMS Mathematical Research Community in Network Science. I cannot help but indulge myself in thanking in particular Leigh Metcalf, Franklin Kenter, Jeremy Kun, Feng 'Bill' Shi, Aaron Clauset, David Kempe and Mason Porter; I feel that interacting with each of you has helped me grow as a mathematical researcher and this work would not be the same without you.

And most importantly, I want to give a special thanks to my advisor, Jonathan Rubin. I am greatly appeaciative of your kindness and patience during my time as a graduate student, since my growth as a mathematical researcher happened in rather large bursts and sometimes it would take some time to see real progress. Furthermore, I do not think I could enumerate all of the ways you have been so supportive of me pursuing my mathematical interests throughout the graduate school experience. I vividly remember when you told me one day in my second year of graduate school that I would be your *networks guy*. It is probably safe to say that at that time neither of us envisioned how much the contents of this thesis would deviate from our initial plans. When I first met you, I probably should have warned you that my high school mathematics teacher told me that he would need a cryptologist to decypher my mathematical writings. And after five years of experiences with one another, I feel very comfortable saying that I can value all of the time and effort you have put in to help me improve my acumen as a mathematical expositor. The quality of this work is a testament to your insightful guidance and talent in conveying and motivating intricate mathematical thought.

I would like to give recognition to my friends and family for their encouragement in completing this piece. My parents have been tremendously supportive of me in my academic endeavors and this work would certainly not be possible without them. I can certainly thank my dad for the long walks on the weekends where he would teach me do mental math problems. Without your help, I would not have possessed a backpack of melted Kit Kat bars that I accumulated from many, many wins in mental math competitions in the third grade. While I definitely lost interest in the Kit Kat bars, the time you have spent with me definitely sparked my interest in the aesthetics of mathematics.

Lastly, I would like to thank my wife for supporting me in completing this work. All your compassion and consideration, there are no words to reciprocate. Lines of numbers and proofs fill my desk; your smile fills my heart.

CHALLENGES IN RANDOM GRAPH MODELS WITH DEGREE HETEROGENEITY: EXISTENCE, ENUMERATION AND ASYMPTOTICS OF THE SPECTRAL RADIUS

David E. Burstein, PhD

University of Pittsburgh, 2016

In order to understand how the network structure impacts the underlying dynamics, we seek an assortment of methods for efficiently constructing graphs of interest that resemble their empirically observed counterparts. Since many real world networks obey degree heterogeneity, where different nodes have varying numbers of connections, we consider some challenges in constructing random graphs that emulate the property. Initially we focus on the Uniform Model, where we would like to uniformly sample from all graphs that realize a given bi-degree sequence. We provide easy to implement, sufficient criteria to guarantee that a bi-degree sequence corresponds to a graph. Consequently, we construct novel results regarding asymptotics of the number of graphs that realize a given degree sequence, where knowledge of the aforementioned enumeration result will assist us in constructing realizations from the Uniform Model. Finally, we consider another random directed graph model that exhibits degree heterogeneity, the Chung-Lu random graph model and prove concentration results regarding the dominating eigenvalue of the corresponding adjacency matrix. We extend our analysis to a more generalized model that allows for intricate community structure and demonstrate the impact of the community structure in networks with Kuramoto and SIS epidemiological dynamics.

Keywords: degree sequence, directed graph, Chung-Lu, random graphs, contingency table, digraph.

TABLE OF CONTENTS

LIST OF FIGURES

1.0 INTRODUCTION

Starting in 2006, medical professionals witnessed a dramatic increase in tuberculosis throughout a community in British Columbia. In order to better understand how the pathogen spread, Gardy et al. [40], used a social network analysis questionairre to identify connections between individuals in the epidemiological network. The resulting set of connections, or edges, between people (nodes), formed a graph corresponding to the network. While empirical data provides much insight as to how tubercuolosis could spread throughout a particular network, empirical data alone does not give us the capability to generalize results concerning the spread of a pathogen to other networks. Instead as applied mathematicians, we want to construct a 'random graph model', where for some choice of parameters, we attain a family of graphs similar to our empirical data and prove that the dynamics (the spread of the pathogen) are essentially the same for all graphs in the collection.

Unsurprisingly, the utility of random graph models extends far beyond epidemiological networks. For example in biological neural networks, pathological amounts of synchronous spiking could be indicative of schizophrenia, Parkinson's or Alzheimer's disease [80]. Consequently using random graph models, we would like to identify families of graphs that promote or abate the likelihood of pathological synchrony. Analogously in ecological networks, where we model populations of species with a system of Lotke-Volterra differential equations, we seek families of graphs that promote or diminish the likelihood of a mass extinction event [74].

We can extend the discussion to genetic networks. Mathematically, we often model genes with boolean dynamics (having an 'on' state or an 'off' state where the state of a gene depends on the states of its 'neighbors'). From a dynamical systems perspective, we want to identify network structures (families of graphs) that are resilient to changes in initial conditions, as many researchers hypothesize that dynamical instability in genetic networks could play a fundamental role in the occurence of cancer [67].

In addition to biological networks, we can consider computer networks, in particular internet routing networks. In the autonomous system internet routing network, autonomous systems receive path announcements through other nodes in the network. Subsequently, autonomous systems send messages to other autonomous systems by selecting an announced path of intermediary autonomous systems to relay the message. Perhaps surprisingly, an autonomous system can announce false paths in order to incentivize other nodes in the network to send messages through the dishonest autonomous system. Even though an autonomous system could announce any false path, to avoid detection, an autonomous system may only consider announcements consisting of a small subset of such paths. Hence, we would want to employ random graph models to identify families of graphs that discourage dishonest behavior [17].

In all of the aforementioned applications, we can define a dynamical process on the graph where the state of the node x depends on the states of the neighbors of x, the nodes that share an edge with x. Furthermore for many dynamical processes, a node x may receive input from a group of nodes, while the state of node x could conceivably affect an entirely different collection of nodes. Naturally, for such a problem we could describe our network as a directed graph where an incoming edge identifies a node that influences $x's$ dynamical state and an outgoing edge identifies the nodes that x can directly influence. Prior research has demonstrated that the bi-degree sequence, a list containing the number of incoming and outgoing edges of each node, can have a signicant impact on the dynamics of the network [64, 71, 39, 72, 63]. Since in real world networks the number of connections from a given node can vary considerably throughout the network, we seek a random graph model that exhibits this property. This serves as motivation for the Uniform Model, where our input consists of the bi-degree sequence and our output is a randomly constructed graph that satisfies that list. Rigorously, we define the Uniform Model as follows.

Random Graph Model 1 (Uniform Model)**.** *Consider the set of all distinct graphs that realize a given bi-degree sequence,* $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{Z}^{N \times 2}$*, a list of the number of incoming and outgoing edges for each of the* N *nodes in the graph. Denote this set of graphs by* $\mathbf{G}$*.*

We define the Uniform Model such that for each graph in the set $\mathbf{G}$*, we randomly choose a graph in* $\mathbf{G}$*, where the probability we select* G_1 *equals the probability we select* G_2 *for any* $G_1, G_2 \in \mathbf{G}$ *.*

Now to generate an assortment of interesting graphs with the Uniform Model we only require a bi-degree sequence. In practice to assess the impact of the degree sequence on the dynamics of the network, we would like to randomly generate many bi-degree sequences and perform numerical simulations on the graphs constructed from the Uniform Model. To this end, suppose we want to randomly construct a bi-degree sequence $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{Z}^{N \times 2}$ where the entries of $\mathbf{a}$ and $\mathbf{b}$ are bounded above by M, below by m and fix the sum of the entries in $\mathbf{a}$ and $\mathbf{b}$ to be S. Then in the spirit of [53], we can randomly sample from all bi-degree sequences that satisfy these properties by first considering an (arbitrary) initial bi-degree sequence that satisfies these constraints; for simplicity, we can choose a degree sequence where all the entries in $\mathbf{a}$ and $\mathbf{b}$ are close to the average entry $\frac{S}{N}$. Then, we choose a random pair of entries in $\mathbf{a}$; we add 1 to the first number in the pair and subtract 1 from the second number in the pair provided that we do not violate the afformentioned presupposed bounds on the degree sequence. If implementing this change would violate our constraints, we do not change the values in the randomly chosen pair. Repeating this step a predetermined number of times for both $\mathbf{a}$ and $\mathbf{b}$ allows us to construct non-trivial degree sequences that satisfies the desired constraints.

Unfortunately, not all bi-degree sequences actually correspond to a graph and we would like to implement a method that would enable us to construct degree sequences that do correspond to a graph; we call such degree sequences *graphic*. This leads us to the following problem.

Problem 1. *Suppose we randomly constructed a degree sequence* $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{Z}^{N \times 2}$ *where the entries of* $\mathbf{a}$ *and* $\mathbf{b}$ *are bounded above by* M*, below by* m *and fix the sum of the entries in* $\mathbf{a}$ *and* $\mathbf{b}$ *to be* S*. Is this randomly constructed degree sequence* $\mathbf{d}$ *graphic? Alternatively, are we guaranteed existence of a graph for all degree sequences that satisfy these constraints?*

At this juncture, we would like to emphasize that even though we motivated Problem 1 by considering a particular method for randomly generating bi-degree sequences (randomly adding and subtracting 1 from the entries in the bi-degree sequence), Problem 1 is in fact relevant to a diverse array of techniques for randomly constructing these bi-degree sequences. More succinctly, we would like to guarantee that our output will yield a graphic bi-degree when utilizing a particular method for randomly constructing bi-degree sequences. For a specific choice of $\mathbf{d}$ we could partially answer Problem 1 by appealing to a classic result, the Gale-Ryser Theorem, which provides criteria guaranteeing the existence of a solution. We explicitly state the Gale-Ryser Theorem below.

Theorem 1. *(Gale-Ryser/Fulkerson [38, 75, 37, 24],[3]) Consider a bidegree sequence* $\mathbf{d} = (\mathbf{a}, \mathbf{b})$ *where the* a_i *are nonincreasing.* $\mathbf{d}$ *is graphic with loops (where we allow a node to have an edge that connects to itself) if and only if*

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i \tag{1.1}$$

and for all $j \in [1..N-1]$,

$$\sum_{i=1}^{n} \min(b_i, j) \geq \sum_{i=1}^{j} a_i.$$

Similarly, $\mathbf{d}$ *is graphic if and only if (1.1) holds and* $\forall j \in \{1, 2, ..., N-1\}$,

$$\sum_{i=1}^{j} \min(b_i, j-1) + \sum_{i=j+1}^{n} \min(b_i, j) \geq \sum_{i=1}^{j} a_i.$$

Foremost, we stress that while the Gale-Ryser Theorem provides a useful check for graphicality for a given degree sequence, in its current form, we cannot easily (or efficiently) adapt the Gale-Ryser Theorem to solve Problem 1 where we seek a guarantee that our output will always yield a graphic degree sequence. As a side node, more recent ammendations of the Gale-Ryser/Fulkerson Theorem do exist and can be found, for example, in Berger [10] and Miller [57]. Miller capitalizes on the discrete "concavity" in j of the functions on the left and right hand sides of the Gale-Ryser inequalities to derive the stronger result. Analogously, we will also exploit the "concavity" in the inequalities to answer Problem 1, which will enable us to construct flexible conditions for easily verifying that our output will always yield a graphic degree sequence.

The results of several past works [82, 2, 19] address a similar problem by providing sufficient conditions for graphicality in terms of the maximum and minimum values in a bidegree sequence, where $\lfloor x \rfloor$ is defined as the integer floor of x.

Theorem 2. *(Zverovich and Zverovich, Alon et al., and Cairns et al.) Consider a bidegree sequence* $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{N}_0^{(N,2)}$ *where* $\mathbf{a} = \mathbf{b}$, $m = \min \mathbf{d}$ *and* $M = \max \mathbf{d}$. *If* $\left\lfloor \frac{(m+M)}{4}^2 \right\rfloor \leq mN$, *then* $\mathbf{d}$ *is graphic with loops.*

Theorem 2 is helpful in the sense that it provides a simple criterion for determining whether there exists a graph corresponding to a given bidegree sequence. Unfortunately, there are in fact bidegree sequences (where $\sum a_i = \sum b_i$) that do not satisfy the conditions and are still graphic, including of course graphic sequences with $\mathbf{a} \neq \mathbf{b}$.

By incorporating an additional quantity, the total number of incoming and outgoing edges of the nodes in a graph, in Chapter 2 we will prove a generalization of Theorem 2, which will provide more flexible criteria for guaranteeing existence even when $\mathbf{a} = \mathbf{b}$ and provide a quick solution to Problem 1.

Historically speaking, Theorem 2 was relevant to the problem of showing that the likelihood that a degree sequence for an undirected graph can produce a graph vanishes as $N \to \infty$ (and an analogous result holds with respect to directed graphs for a bidegree sequence that have equal in- and out-degree sums and is otherwise unconstrained). The constraint on the maximum of the bidegree sequence in the above theorem suggested that the probability of graphicality would approach zero in this limit, since excessive growth of M (e.g., proportional to N) with increasing N would violate the graphicality condition [33, 5, 6, 62]. Ultimately, a result from Pittel [66] provides a proof of this asymptotic result. In general, identifying families of graphic degree sequences is a nontrivial problem and constructing improved sufficient conditions for graphicality can help ease this difficulty.

We would like to emphasize that we truly have multiple motivations for pursing novel sufficient conditions for graphicality. Firstly, even for a fixed bidegree sequence $\mathbf{d}$, determining whether $\mathbf{d}$ is indeed graphic from the N inequalities in Theorem 1 is conceptually cumbersome. Inspection of a given degree sequence provides little intuition as to whether it is possible to construct a graph that realizes that degree sequence. Additionally, verifying the N inequalities in Theorem 1 directly can also be computationally inefficient. While linear time algorithms exist for verifying the conditions in the Gale-Ryser Theorem [43, 50, 34], if we knew the maximum and minimum of our degree sequence, we would like to invoke a criteria similar to Theorem 2 to have an $O(1)$ check for graphicality as opposed to an $O(N)$ check. Consequently, we aim to strengthen Theorem 2 as we often want to sample *many* different bidegree sequences and hence using a linear time check would be inefficient. For example, generating a large graph using the methods adopted by Kim et al. [49] requires taking a node from the graph and identifying all wirings of its outward edges that can lead to a digraph without multi-edges. To do so, one must check graphicality of the residual bidegree sequence many times; avoiding this step by utilizing the conditions formulated in Chapter 2 that ensure graphicality could help speed up the run time of the code.

After exploring the intricacies of identifying degree sequences that correspond to graphs, to construct such graphs, we seek a method for deriving expressions regarding the probability that two nodes share an edge under the Uniform Model. We claim (and argue shortly) that providing a solution to the following problem will assist us immensely in this regard.

Problem 2. *Given a degree sequence* $\mathbf{d} = (\mathbf{a}, \mathbf{b})$*, how many different graphs realize this degree sequence? Equivalently, how many* $0 - 1$ *binary matrices have rows sums given by* $\mathbf{a}$ *and column sums given by* $\mathbf{b}$*?*

Before proceeding, it is mathematically convenient to represent a (directed) graph as a $0 - 1$ binary matrix where the ijth entry is 1 if there is an edge from node j to node i and 0 otherwise. In this context, we can frame the problem of counting the number of different graphs that correspond to a degree sequence as the number of $0 - 1$ binary matrices with prescribed row and column sums. To demonstrate the connection between counting the number of graphs that realize a given degree sequence and the probability that two nodes share an edge, we present the following example.

Example 1. *Consider* $0 - 1$ *binary matrices that have row sums* $(2, 2, 1)$ *and column sums* $(2, 1, 2)$.

$$\begin{array}{cccc} 2 & 1 & 2 & \\ A_{11} & A_{12} & A_{13} & 2 \\ A_{21} & A_{22} & A_{23} & 2 \\ A_{31} & A_{32} & A_{33} & 1 \end{array}$$

What percentage of such matrices have $A_{31} = 1$*?*

To answer this question, let's rewrite the above matrix substituting the value $A_{31} = 1$*.*

$$\begin{array}{cccc} 2 & 1 & 2 & \\ A_{11} & A_{12} & A_{13} & 2 \\ A_{21} & A_{22} & A_{23} & 2 \\ 1 & A_{32} & A_{33} & 1 \end{array}$$

Notice that the last row sum must equal 1 *so that forces* $A_{32} = A_{33} = 0$*. In addition, we have the constraint that* $A_{11} + A_{21} + 1 = 2$*. Hence, we want to count the number of* $0-1$ *binary matrices of the form,*

$$\begin{array}{cccc} 1 & 1 & 2 & \\ A_{11} & A_{12} & A_{13} & 2 \\ A_{21} & A_{22} & A_{23} & 2 \end{array}$$

Consequently, under the Uniform Model the likelihood that $A_{31} = 1$ *(there is an edge from node* 1 *going to node* 3*) is precisely the number of* $0-1$ *binary matrices with row sums* $(1,1,2)$ *and column sums* $(2,2)$ *divided by the number of* $0-1$ *binary matrices with row sums* $(2,1,2)$ *and column sums* $(2,2,1)$*.*

In Chapter 3 we provide an asymptotic solution to the counting problem, Problem 2, by exploiting the intuition that for *sparse* networks, two fixed nodes should not have common neighbors with high probability. Such an approach ultimately yields a recursion that we can manipulate to asymptotically estimate the number of graphs that realize a given degree sequence that satisfies certain sparsity constraints.

As hinted in Example 1, counting the number of graphs that realize a given degree sequence is an important step for uniformly generating $0-1$ matrices (contingency tables) with fixed row and column sums. While the methods proposed for (almost) uniformly generating graphs are quite diverse, many (if not most) of the methods proposed previosly can be classified into two categories, Markov Chain Monte Carlo (MCMC) methods and Sequential Importance Sampling (SIS).

The traditional MCMC method involves swapping edges many times to generate an approximately uniformly random sample. The main drawback of this method is that there is an unknown mixing time. Naturally, by knowing precise asymptotics for the number of graphs with a prescribed degree sequence, we can deduce asymptotic probabilities for the likelihood two nodes share an edge. Consequently, we could use these asymptotic probabilities as a criteria to empirically help us determine the mixing time. There are quite a few interesting technical issues for implementing this method and we refer the reader to the literature for details [59, 46, 45, 12, 41, 78]. We also mention that for nicely behaved degree sequences, there are MCMC methods that have provable bounds for almost uniformly sampling graphs with a prescribed degree sequence, but they are also computationally expensive [13].

Alternatively, SIS methods sample the number of graphs with a prescribed degree sequence in a biased way. A large sample of graphs is taken from a biased distribution and a Law of Large Numbers argument is used to construct a new (approximately) uniform distribution based on the output of the biased sampling procedure [16, 25, 32]. The biggest drawback of SIS is that we often do not know how large a sample of graphs we need from our biased distribution to reliably construct the approximately uniform distribution. Indeed, past work has shown that certain constraints on degree sequences may be required to ensure the computational efficiency of SIS methods [8, 15]; without such constraints, an exponentially large sample size may be required to attain meaningful estimates for approximating the uniform distribution [14]. Statistical arguments show that to increase the speed of convergence, we want an initial biased distribution that is quite close to the uniform distribution [4, 15]. Incorporating asymptotically accurate graph enumeration formulas in constructing a biased distribution could conceivably improve the performance of such methods.

Historically, the derivation of asymptotic formulas for the number of $0-1$ binary matrices with fixed row and column sums has a rich history spanning at least as far back as 1958 with Read [70]. Progress in this area has been made by restricting to matrices that are sparse [9], in the sense that row and column sums grow at most as a fractional power of the norm of the matrix or by restricting to matrices that are dense but have limited variation among the row and column sums [7, 20]. As mentioned earlier, we focus on the sparse case. In this setting, McKay [54] developed a formula to count such matrices that is valid in an asymptotic sense in the limit as the number of edges S becomes arbitrarily large, assuming that the maximum row sum or column sum grows as $o(S^{\frac{1}{4}})$. More recently, Greenhill et al. [42] generalized McKay's formula, obtaining a result that holds if the maximum row sum or column sum is $o(S^{\frac{1}{3}})$. Since many real world networks can consist of many thousands of nodes, commonly with connectivities of up to 10 %, we aim to count matrices with row and column sums, corresponding to in- and out-degrees, that exceed $O(S^{\frac{1}{3}})$. Furthermore, since we are asymptotically guaranteed graphicality provided that the maximum degree (row sum or column sum) is bounded by $\sqrt{S}$, as demonstrated by our graphicality results in Chapter 2, we expect that we can extend the asymptotic enumeration results of McKay and Greenhill to allow for the maximum degree to reach $O(S^{\frac{1}{2}-\tau})$ for any $\tau > 0$. As such we dedicate Chapter 3 to this endeavor.

In order to accurately model real world networks we want to prove results pertaining to a diverse assortment of random graph models, not just the Uniform Model. Consequently, we consider another random graph model that also exhibits 'degree heterogeneity', the Chung-Lu random graph model [27, 28].

Random Graph Model 2 (Chung-Lu)**.** *Consider an (expected) bi-degree sequence,* $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{Z}^{N\times 2}$*, for each of the* N *nodes in the graph, where* $\max_i a_i \max_j b_j \leq S = \sum_i a_i = \sum_i b_i$ *Then with independent probability* p_{ij} *we construct an edge from node* j *to node* i*, where* $p_{ij} = \frac{b_j a_i}{S}$

We dedicate Chapter 4 to ascertain how the parameters (the expected degree sequence) of Chung-Lu random graphs can impact certain features in our network. Prior literature has demonstrated how the spectral radius can impact the dynamics in epidemiological, neuronal and genetic networks [64, 39, 67, 80, 35, 63]. Furthermore, while many results regarding the spectral radius pertaining to realizations of undirected graphs exist [29, 30, 52, 21, 60, 65, 51], very few (if any) rigorous results have been proven in context to the distribution of the spectral radii for random directed graphs. Consequently, we prove a conjecture from

Restrepo, Ott and Hunt [73] regarding the limiting distribution for the spectral radii of these adjacency matrices for realizations of the directed Chung-Lu random graph model. Rigorously, under the assumptions that $\max_{i,j} p_{ij} \to 0$ or $\frac{\mathbf{a}\cdot\mathbf{b}}{S} \to \infty$, we show that almost surely the dominating eigenvalue of a random adjacency matrix realization $\rho(A) \to \frac{\mathbf{a}\cdot\mathbf{b}}{S}$.

Subsequently, we study the spectral radius of graph realizations from a more general random graph model, which we call the **Partitioned Chung-Lu** random graph model [22, 48, 81, 79]. In the Partitioned Chung-Lu random graph model for each node x, we define a function $G(x)$ to identify $x's$ corresponding group. We then model the likelihood a node x has an outgoing connection to node y as an independent Bernoulli random variable with probability $\frac{b_x(G(y))a_y(G(x))}{S(G(x),G(y))}$ where $b_x(G(y))$ is the expected number of outgoing edges of node x when the receiving node belongs to group $G(y)$, $a_y(G(x))$ is the expected number of incoming edges of node y where the outgoing edge comes from a node in group $G(x)$ and $S(G(x), G(y))$ is the expected number of edges that start from a node in group $G(x)$ and end in the group $G(y)$. We emphasize that analogous to the Chung-Lu random graph model, for all choices of nodes x and y, $b_x(G(y))$, $a_y(G(x))$ and $S(G(x), G(y))$ are parameters for the Partitioned Chung-Lu random graph model.

Remarkably, even though the generality of the Partitioned Chung-Lu random graph model allows for the generation for networks with exceptionally intricate community structure, the assumption that the edge probabilities in each partition emulates the behavior of a Chung-Lu random graphs enables us to carry over the analysis to this significantly more general case. More specifically, in the case that we partition our adjacency matrix with m submatrices on each row and m submatrices on each column, we argue that asymptotically the dominating eigenvalue of realizations from the directed Partitioned Chung-Lu random graph model equals the dominating eigenvalue of an $m^2 \times m^2$ entry-wise non-negative matrix (which we explicitly state in Chapter 4). To derive this result, consider the first row in our partitioned adjacency matrix. We first 'average' the interaction of the first partitioned row on each of the m partitioned columns and record each of these m entries in our reduced matrix. We then repeat this process for each partitioned row, resulting in an $m^2 \times m^2$ entry-wise non-negative matrix. In the limit, we expect these 'averages' to exemplify the structure of the network partitions remarkably well and hence we will attain convergence.

2.0 SUFFICIENT CONDITIONS FOR GRAPHICALITY

Generating random graphs with various properties is relevant for a wide variety of applications, from modeling neural networks [80] to internet security [1]. To generate an undirected random graph with a fixed number of nodes, it is natural to first select a degree distribution through some process and then to connect the nodes in a way that is consistent with the selected distribution; similarly, a bidegree distribution would be selected if a directed graph were desired. A well known issue with this procedure is that not all degree distributions are graphic; that is, it is easy to write down a sequence of n natural numbers $\{d_i\}$ such that there is no graph with n nodes for which the degree of the ith node is d_i for all i. The aim of this work is to rigorously establish novel, relatively inclusive, easily checked conditions on a bidegree sequence that ensure that it is graphic and hence corresponds to one or more directed graphs. Such conditions can be used as constraints on a degree distribution to ensure that sampling from that distribution will yield a graphic degree sequence or to ease the process of verifying that a (randomly generated) bidegree sequence corresponds to a directed graph.

To start, we briefly review some standard definitions. In doing so, and in the rest of the chapter, we will employ what is known as Hoare-Ramshaw notation for closed sets of integers, namely $[a..b] := \{x \in \mathbb{Z} : a \leq x \leq b\}$ for $a, b \in \mathbb{Z}$. We will also define $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ and $\mathbb{N}_0^{(n,2)} = \{(\mathbf{a}, \mathbf{b}) : \mathbf{a} \in \mathbb{N}_0 \text{ and } \mathbf{b} \in \mathbb{N}_0\}$.

Definition 1. *A bidegree sequence* $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{N}_0^{(n,2)}$ *is graphic if there is a 0-1 binary matrix with 0's on the main diagonal such that the sum of the ith row is* a_i *and ith column is* b_i *for all* $i = [1..n]$*. We say a bidegree sequence* $\mathbf{d} \in \mathbb{N}_0^{(n,2)}$ *is graphic with loops if there is a 0-1 binary matrix such that the sum of the ith row is* a_i *and ith column is* b_i*. We call* $\mathbf{a}$

our in-degree sequence and $\mathbf{b}$ *our out-degree sequence.*

Note that when it exists, the 0-1 binary matrix in Definition 1 arises naturally as the adjacency matrix for the digraph with degrees given by $\mathbf{a}, \mathbf{b}$. In this matrix, the (i, j) element is 1 if the digraph includes an edge from node j to node i and a 0 if it does not. To distinguish graphicality for digraphs from that for graphs, one might refer to the statement of Definition 1 as defining what it means for $\mathbf{d}$ to be digraphic. For simplicity, we shorten this to graphic since we do not focus on undirected graphs in this paper.

As a final note, all of the results in this chapter immediately extend to graphicality results for bipartite graphs, since every bipartite graph can be represented as a $0-1$ rectangular binary matrix. We can extend any rectangular binary matrix as a square binary matrix by adding rows (or columns) of $0's$. Since there is a one-to-one correspondence between digraphs (with loops) and square $0-1$ binary matrices, any sufficient conditions that guarantee existence for digraphs carry over to bipartite graphs as well.

2.1 EXPLOITING CONCAVITY TO CONSTRUCT NOVEL SUFFICIENT CONDITIONS

To start, we prove the following Theorem, which considers the maximum of the in-degree and the maximum of the out-degree as two separate parameters.

Theorem 3. *Consider a bidegree sequence* $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{N}_0^{(n,2)}$ *where the entries of* $\mathbf{a}$ *are arranged in non-increasing order and assume that* $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i := n\bar{c}$ *where* $\bar{c}$ *is the average degree. If* $\max a_i = M_a$ *and* $\max b_i = M_b$*, where* $M_a M_b \leq n\bar{c} + 1$*, then* $\mathbf{d}$ *is graphic with loops. In particular, in the special case where* $M_a = M_b$*, if* $\max \mathbf{d} \leq \left\lfloor \sqrt{n\bar{c}+1} \right\rfloor$*, then* $\mathbf{d}$ *is graphic with loops.*

We will prepare for the proof of the theorem with certain preliminary results. Before doing so, we want to point out that that the adjustment of the bounds needed to ensure graphicality Theorem 4 rather than graphicality with loops Theorem 3 is quite small. This should not be surprising as graphicality requires that the adjacency matrix have 0's on the main diagonal. Since this restriction only affects n of the n^2 entries in our adjacency matrix, it should have negligible impact in the limit of large n. This concept appears again later when we consider different sufficient criteria. Thus, in both instances, after we prove a sufficient condition to ensure graphicality with loops, we will make a slight alteration to our sufficient condition and show that the new version guarantees the (slightly) stronger condition of graphicality.

Now, in the sufficient criteria in Theorem 2, for simplicity suppose that $\left\lfloor \frac{(m+M)^2}{4} \right\rfloor = \frac{(m+M)^2}{4}$, such that $\frac{(m+M)^2}{4} \leq mn$ implies $M \leq \sqrt{4mn} - m \leq \sqrt{4mn}$. We conclude that if $\bar{c} > 4m$, then Theorem 3 (with $M_a = M_b$), provides a more flexible criterion for graphicality than that given by Theorem 2.

We also wish to differentiate Theorem 3 from the constraint provided by Chung and Lu [27]. In their work, the probability of having an outgoing edge from node j to node i is given by a Bernoulli random variable p_{ij}, independent across choices of i, j, such that $p_{ij} = \frac{a_i b_j}{n\bar{c}}$ where a_i is the in-degree of node i and b_j is the out-degree of node j. Consequently, they require that $M_a M_b \leq n\bar{c}$ in order to ensure that the probabilities do not exceed 1. It is not at all obvious that this bound should translate into a sufficient condition for graphicality, and it can in fact be awkward for the Chung-Lu algorithm. Specifically, if $M_a M_b = n\bar{c}$, and there exists a node i such that $a_i = M_a$, and a node j such that $b_j = M_b$, then according to the Chung-Lu algorithm, the probability of constructing an edge between node i and node j is 1, which is not a natural choice [77].

To begin the analysis, consider all bidegree sequences in $\mathbb{N}_0^{(n,2)}$ with maximum in-degree M_a, with maximum out-degree M_b, and with average degree $\bar{c}$, such that $n\bar{c}$ is the sum of the in-degrees and also the sum of the out-degrees. To prove Theorem 3, we want to construct the worst possible scenario; that is, we want to identify the in-degree vector that for *each and every* j maximizes $\sum_{i=1}^{j} a_j$, and the out-degree vector that for *each and every* $j \in [1..n-1]$, minimizes $\mathbf{F}(j, \mathbf{b}) := \sum_{i=1}^{n} \min(b_i, j)$ (Recall Theorem 1. Once we verify that the n inequalities still hold under this worst case scenario, we have consequently proved the theorem. Since identifying the minimizer of $\mathbf{F}(j, \mathbf{b})$ is rather technical, we prove the result in the following Lemma and Corollary for clarity; Lemma 1 also follows from Lemma 2.3 in [57] with $a_k = \Psi(k) - \Phi(k)$ as defined below. Notice, however, that $\mathbf{F}(j, \mathbf{b})$ in Corollary

1 is *not* defined as $\sum_{i=1}^{n}\min(b_i, j)$ and for completeness we will show later in the proof of Theorem 3 that indeed

$$\sum_{i=1}^{n}\min(b_i, j) = \sum_{i=1}^{j}\#(b_z : b_z \geq i, 1 \leq z \leq n). \tag{2.1}$$

Lemma 1. *Let $\Phi : \mathbb{N} \to \mathbb{N}$, $\Psi : \mathbb{N} \to \mathbb{N}$, where Ψ is a concave function; that is, $\nabla\Psi(j) = \Psi(j) - \Psi(j-1)$ is non-increasing in j. Let $\gamma \in \mathbb{N}$. If $\nabla\Phi(j) = \Phi(j) - \Phi(j-1) = \gamma$ or $\nabla\Phi(j) = \gamma - 1$ for all $j \in [\alpha..\beta]$, $\Phi(\alpha) \leq \Psi(\alpha)$ and $\Phi(\beta) \leq \Psi(\beta)$, then $\Phi(j) \leq \Psi(j)$, for all $j \in [\alpha + 1..\beta - 1]$.*

Proof. Suppose that there exists a first contradiction such that $\Phi(k) > \Psi(k)$, for some k. This implies that $\nabla\Phi(k) > \nabla\Psi(k)$ as $\Phi(k-1) \leq \Psi(k-1)$. But since by assumption and concavity, $\nabla\Psi(j) \leq \nabla\Psi(k) \leq \nabla\Phi(k) - 1 \leq \min(\nabla\Phi(j))$ for all $j > k$, this implies that $\Phi(j) > \Psi(j)$ for all $j > k$. Since we assumed that $\Phi(\beta) \leq \Psi(\beta)$, we have arrived at a contradiction. □

Corollary 1. *For $\mathbf{b} \in \mathbb{N}_0^n$, let $\mathbf{F}(j, \mathbf{b}) = \sum_{i=1}^{j}\#(b_z : b_z \geq i, 1 \leq z \leq n)$. Fix $M \in \mathbb{N}$ and define the set B_M of out-degree vectors as $B_M := \{\mathbf{b} \in \mathbb{N}_0^n : \sum_{i=1}^{n} b_i = n\bar{c}, \max_i b_i \leq M$, and $M \leq n\bar{c}\}$. Choose $k \in \mathbb{N}$ with $k \leq n$ such that $kM \leq n\bar{c}$ and $(k+1)M > n\bar{c}$. Define $\mathbf{b}^*$ as $b_1^* = \ldots = b_k^* = M$, $b_{k+1}^* = n\bar{c} - kM$ and $b_l^* = 0$ for all $l > k+1$. Then under these assumptions, for every $\mathbf{b} \in B_M$, $\mathbf{F}(j, \mathbf{b}^*) \leq \mathbf{F}(j, \mathbf{b})$ for each and every $j \in [1..n]$.*

Proof. Fix $M \in \mathbb{N}$. Note that $\mathbf{F}(j, \mathbf{b}) = \sum_{i=1}^{j}\#(b_z : b_z \geq i, 1 \leq z \leq n)$ is concave in j and $\mathbf{F}(M, \mathbf{b}) = n\bar{c}$ for all $\mathbf{b} \in B_M$. For $\mathbf{b}^*$ as defined in the statement of the Corollary, it follows that $\mathbf{F}(1, \mathbf{b}^*) \leq \mathbf{F}(1, \mathbf{b})$ for all $\mathbf{b} \in B_M$. Note that there at most $k+1$ positive entries in the out-degree sequence $\mathbf{b}^*$ and k of them are identical. Consequently, for $j \leq M$, there are k or $k+1$ entries in $\mathbf{b}^*$ with entries that equal or exceed j and by definition $\nabla\mathbf{F}(j, \mathbf{b}^*) = \#(b_z : b_z \geq j, 1 \leq z \leq n) \in \{k, k+1\}$. Hence, applying Lemma 1 yields the desired result. □

At this juncture, we now can prove **Theorem 3**.

Proof. Consider a bidegree sequence $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{N}_0^{(n,2)}$ satisfying the assumptions of Theorem 3. Let us use the out-degrees to construct an n by n matrix consisting only of zeros and ones (a Ferrers diagram). For the kth column, starting with the first row, we write down a 1. We continue writing 1's until the column sums to b_k and let the remaining entries in the column be zero. Denote the kth row sum as Q_k. It follows algebraically that $\sum_{i=1}^{n} \min(b_i, j) = \sum_{i=1}^{j} \sum_{k=1}^{n} \mathbf{1}_{(b_k \geq i)} = \sum_{i=1}^{j} Q_i = \sum_{i=1}^{j} \#(b_z : b_z \geq i, 1 \leq z \leq n)$.

Graphicality is trivial if $M_a = 1$ as long as $\sum_i a_i = \sum_i b_i$. For $M_a > 1$, we have proven that the minimizer has out-degree sequence $\mathbf{b}^*$ such that $b_1^* = ... = b_{M_a-1}^* = M_b$ (as $M_a M_b \leq n\bar{c} + 1$). Now, if $M_a M_b \leq n\bar{c}$, then $b_{M_a}^* = M_b$ as well. On the other hand, if $M_a M_b = n\bar{c} + 1$, then $b_{M_a}^* = n\bar{c} - (M_a - 1)M_b = M_b - 1$. Hence, we are assured that $b_{M_a}^* \geq M_b - 1$.

Consequently, for $j \leq M_b - 1$, $\sum_{i=1}^{j} a_i \leq jM_a$, as $\max a_i \leq M_a$, and furthermore, $jM_a = \sum_{i=1}^{j} \#(b_k^* \geq i) \leq \sum_{i=1}^{n} \min(b_i, j)$. Meanwhile, for all $j \geq M_b$, $\sum_{i=1}^{n} \min(b_i, j) = n\bar{c}$. Hence, by Theorem 1, the result is proved. □

The sufficient condition in Theorem 3 is the best we can do without knowing more information regarding our degree sequence, as illustrated in the following counterexample.

Counter Example 1. *There exists a degree sequence* $\mathbf{d}$ *such that* $\sum a_i = \sum b_i = n\bar{c}$, $M_a M_b = n\bar{c} + 2$, *and* $\mathbf{d}$ *is not graphic with loops.*

Proof. Consider a degree sequence with $M_a \geq 2, M_b > 2$ where $b_1 = ... = b_{M_a-1} = M_b$ and $b_{M_a} = n\bar{c} - (M_a - 1)M_b = M_b - 2$. Furthermore let $a_1 = ... = a_{M_b-1} = M_a$. Then it follows that this degree sequence is not graphic as,

$$\sum_{i=1}^{M_b-1} a_i = M_a(M_b - 1) > \sum_{i=1}^{M_b-1} \#(b_k \geq i) = M_a(M_b - 2) + (M_a - 1) = M_a(M_b - 1) - 1.$$

□

With a subtle but natural observation we can generalize Theorem 3 to the case where we prohibit loops and the bound will be remarkably similar.

Theorem 4. *Consider a bidegree sequence* $\mathbf{d} \in \mathbb{N}_0^{(n,2)}$ *where* $\sum a_i = \sum b_i = n\bar{c}$*. If* $\max a_i \leq M_a$ *and* $\max b_i \leq M_b$*, where* $(M_a+1)M_b \leq n\bar{c}$*, then* $\mathbf{d}$ *is graphic. In particular, if* $\max \mathbf{d} = M_a = M_b \leq \sqrt{\frac{1}{4}+n\bar{c}} - \frac{1}{2}$*, then* $\mathbf{d}$ *is graphic.*

Proof. First, we show that for $j \leq M_b$, the jth inequality from the Gale-Ryser Theorem holds. We have $\sum_{i=1}^{j} a_i \leq jM_a$ and

$$jM_a = j(M_a+1) - j \leq_* \sum_{i=1}^{n} \min(b_i, j) - j \leq \sum_{i=1}^{j} \min(b_i, j-1) + \sum_{i=j+1}^{n} \min(b_i, j).$$

The starred inequality follows from applying Lemma 1 to minimize the sum $\sum_{i=1}^{n} \min(b_i, j)$ with respect to the constraint that $\max(b_i) \leq M_b$, where

$$\sum_{i=1}^{n} \min(b_i, j) = \sum_{i=1}^{j} \sum_{k=1}^{n} \mathbf{1}_{(b_k \geq i)} = \sum_{i=1}^{j} \#(b_z : b_z \geq i, 1 \leq z \leq n),$$

$M_b(M_a+1) \leq n\bar{c}$, and $\sum_{i=1}^{n} b_i = n\bar{c}$. For the minimizing sequence thus obtained, $b_1^* = ... = b_{M_a+1}^* = M_b$, as $M_b(M_a+1) \leq n\bar{c}$, and hence

$$\sum_{i=1}^{n} \min(b_i^*, j) = j(M_a+1).$$

For $j \geq M_b + 1$, we can eliminate the minimum functions, as now $j - 1 \geq M_b \geq b_i$ for all i. Thus,

$$\sum_{i=1}^{j} \min(b_i, j-1) + \sum_{i=j+1}^{n} \min(b_i, j) = \sum_{i=1}^{n} b_i,$$

and $\sum_{i=1}^{n} b_i \geq \sum_{i=1}^{j} a_i$ for $j \leq n$ as the a_i's are nonnegative and by assumption $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i$.

In the special case where $M_a = M_b$, it follows that $M_a = \left\lfloor \sqrt{\frac{1}{4}+n\bar{c}} - \frac{1}{2} \right\rfloor$, as this quantity is the largest integer that satisfies the inequality $M(M+1) \leq n\bar{c}$.

□

For large graphs, Theorems 3 and 4 provide bounds that ensure graphicality of a bidegree sequence while allowing for a relatively large maximal degree. However, for many graphs we also have information about a lower bound on in- and out-degrees. Consequently, we aim to prove two types of extensions. In one extension, Theorem 5, we assume that there is a nonzero minimum degree, which in turn enables us to construct a more flexible sufficient condition on the maximum degree to guarantee graphicality. The other type of extension, given in Corollary 5, also exploits the working assumption of a minimum degree in order to allow a small set of exceptional degrees to exceed the bound on the maximum proposed in Theorem 3 while maintaining graphicality. We explore these issues in the following section.

2.2 REFINING OUR SUFFICIENT CONDITIONS WITH THE MINIMUM DEGREE

To prove the desired extensions, it is convenient to not have to manipulate two seperate values for the maximum of the in-degree sequence M_a and the maximum of the out-degree sequence M_b, Henceforth, we drop the notation M_a and M_b and efer to the maximum value of the bidegree sequence as M, given by

$$M = \max_i \{\max_i a_i, \max_i b_i\}.$$

Furthermore, throughout this section we will assume that $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i$, based on the necessity of this equality for graphicality.

Corollary 2. *Suppose that a bidegree sequence* $\mathbf{d} \in \mathbb{N}_0^{(n,2)}$ *has a maximum value* $M < n$ *and for the associated in-degree sequence,* $\#(a_i = M) = k$*, where* $M \leq k$*. Then* $\mathbf{d}$ *is graphic with loops. More generally, if for some* $k \in \mathbb{N}$*, both* $M \leq k$ *and* $Mk \leq n\bar{c}$ *hold, then* $\mathbf{d}$ *is graphic with loops.*

Proof. It follows by assumption that $M^2 \leq Mk \leq n\bar{c}$ and hence Theorem 3 applies. □

An application of Corollary 2 provides us with a powerful check for graphicality with loops. Indeed, suppose we have verified the first k inequalities of the Gale-Ryser theorem where the maximum is large ($M >> \sqrt{n\bar{c}}$). We can then look at the residual degree sequence where the residual maximum is much more friendly and construct a linear upper bound for the remaining inequalities based on the new maximum of the residual degree sequence to verify whether the remaining $n-k$ inequalities hold.

Before we move on to prove Theorem 5, we make an adjustment to Corollary 2 to handle graphicality without loops.

Corollary 3. *Suppose that a bidegree sequence* $\mathbf{d} \in \mathbb{N}_0^{(n,2)}$ *has a maximum value* $M < n$ *and for the associated in-degree sequence,* $\#(a_i = M) = k$*, where* $M < k$*. Then* $\mathbf{d}$ *is graphic. More generally, if there exist* $k, M \in \mathbb{N}$*, such that* $M < k$ *and* $Mk \leq n\bar{c}$*, then* $\mathbf{d}$ *is graphic.*

Proof. The proof is analogous to the prior corollary, since the assumptions give $M(M+1) \leq Mk \leq n\bar{c}$ and application of Theorem 4 completes the proof. □

We are now ready to prove the following sufficient condition on graphicality, and later we show that it is an asymptotically sharp refinement over the condition proven by Zverovich and Zverovich [82].

Theorem 5. *Consider a bidegree sequence* $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{N}_0^{(n,2)}$ *where* $\sum a_i = \sum b_i = n\bar{c}$ *and* $\min \mathbf{d} = m \in [1..n]$*. Define*

$$k_* = m + \sqrt{m^2 + n(\bar{c} - 2m)} \tag{2.2}$$

and let $k = \lceil k_* \rceil$ *if* k_* *is real and* $k = 1$ *otherwise. If*

$$M := \max \mathbf{d} \leq \min(\left\lfloor n\frac{\bar{c} - m}{k} + m \right\rfloor, n), \tag{2.3}$$

then $\mathbf{d}$ *is graphic with loops.*

Proof. To start, suppose that a bidegree sequence has maximal degree M given by (2.3) with k_*, k as defined in the statement of the theorem. Note that $n\bar{c} - M \leq Mk + (n-(k+1))m \leq n\bar{c} - m$. We can thus consider the vector $\mathbf{b}^*$ such that $b_1^* = b_2^* = ... = b_k^* = M$, $b_{k+1}^* = r$ and all other $b_i^* = m$, where we choose the remainder r such that $kM + r + (n-(k+1))m = n\bar{c}$ and thus $m \leq r \leq M$. Recall that $\mathbf{F}(j, \mathbf{b}) = \sum_{i=1}^n \min(b_i, j)$ and that we have an alternative representation of $\mathbf{F}(j, \mathbf{b})$ from (2.1). Since $\mathbf{F}(m, \mathbf{b}) = nm$ for all $\mathbf{b}$ with minimum m, we can apply Lemma 1 to show that $\mathbf{b}^*$ is a minimizer of $\mathbf{F}$.

At this stage, we assume that k_* is real, and we would like to show that the first k Gale-Ryser inequalities hold. In fact, because of the nonzero minimum m, the first m Gale-Ryser inequalities are trivially satisfied since $M \leq n$; in particular, in the special case of $m = n$, Theorem 5 is true. So, the only case we need to consider here is the case when $m < k$ and $m < n$, which we henceforth assume. As previously, $\sum_{i=1}^j a_i \leq jM =: V(j)$. Since $V(j)$ is linear and $W(j) := \sum_{i=1}^n \min(b_i, j)$ is concave, by Lemma 1, to verify the first k inequalities of the Gale-Ryser Theorem, it suffices to show that $V(k) \leq W(k)$. Therefore, we seek to verify that the kth inequality holds for our minimizing vector $\mathbf{b}^*$.

The definition of r implies that the following two equivalent equations both hold:

$$kM + r - m + (n-k)m = n\bar{c} \iff kM + r = n\bar{c} - (n-k-1)m. \tag{2.4}$$

Using $r \geq m$ in (2.4), it follows that

$$\sum_{i=1}^k a_i \leq kM \leq n\bar{c} - (n-k)m. \tag{2.5}$$

Furthermore, for $m < k$ and $m < n$,

$$nm + k(k-m) \leq \sum_{i=1}^n \min(b_i^*, k) \leq \sum_{i=1}^n \min(b_i, k), \tag{2.6}$$

since the middle quantity is $k^2 + \min(r, k) + (n-(k+1))m$ and $\min(r, k) \geq m$.

Combining (2.5) and (2.6) implies that the first k Gale-Ryser inequalities will be guaranteed to hold as long as $n\bar{c} - (n-k)m \leq nm + k(k-m)$ or, equivalently, as long as

$$R(k) = k^2 - 2mk + 2nm - n\bar{c} \geq 0.$$

Note that $R(k) \geq 0$ for $k \geq k_*$ where k_* as defined in (2.2), which we have assumed for now to be real, is the larger root of $R(k)$. Unfortunately, k_* does not have to be a natural number. But for $k = \lceil k_* \rceil = k_* + z$, where $z := k - k_* \in [0, 1)$, it follows that $R(k) \geq 0$.

We have now established that under this choice of M, the first k Gale-Ryser inequalities hold, where k may be equal to 1 or k_* depending on whether or not k_* is real. We no longer assume that $k \neq 1$, and we next proceed to show that the remaining inequalities hold as well. Assume that we have a remainder $b^*_{k+1} = r > m$, and we wish to verify the $(k+1)$st inequality for our minimizing vector. We will construct another polynomial, $S(\cdot)$, such that if the polynomial is nonnegative when evaluated at $(k+1)$, then the $(k+1)$st inequality in the Gale-Ryser Theorem holds. Furthermore we will show that for our choice of k, $S(u) \geq 0$ for $u \geq k$.

It follows from equation (2.4) that $kM + r = n\bar{c} + m - (n-k)m$ and we would like to find a condition on k that ensures that $n\bar{c} + m - (n-k)m \leq nm + k(k+1-m) + 1 \leq \mathbf{F}(k+1, \mathbf{b}^*)$, where the $+1$ in the middle quantity is a lower bound on r. We therefore define

$$S(k) = k^2 + k(1 - 2m) + (2n - 1)m - n\bar{c} + 1,$$

with largest root

$$k_{**} = m - \frac{1}{2} + \sqrt{m^2 + n\bar{c} - 2nm - \frac{3}{4}}$$

(if the roots are positive), and by an analogous argument to that used for $R(\cdot)$, it follows that for all $u \geq k_{**}$, $S(u) \geq 0$. By noting that $k \geq k_* > k_{**}$ (if k_* is real), we have shown that

$$\sum_{i=1}^{k+1} a_i \leq F(k+1, \mathbf{b}).$$

To finish off the proof, it remains to verify the $\{k+2, k+3, ..., n\}$ inequalities. Define $\delta = 1$ if $r > m$ and 0 otherwise. Since we have shown that $\sum_{i=1}^{k+\delta} a_i \leq \sum_{i=1}^{k+\delta} \min(b^*_i, k+\delta) \leq \mathbf{F}(k+\delta, \mathbf{b})$ and we know that $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} \min(b^*_i, n) = \mathbf{F}(n, \mathbf{b})$, Lemma 1 guarantees that for all j such that $k + \delta \leq j \leq n$, $\sum_{i=1}^{j} a_i \leq \sum_{i=1}^{n} \min(b^*_i, j) \leq \mathbf{F}(j, \mathbf{b})$. Thus, the proof is complete. $\square$

Now we state the analogous result to show that a degree sequence is graphic. We also provide a sketch of the proof, which follows similarly to the proof of Theorem 5, and leave it to the reader to fill in the missing details.

Theorem 6. *Consider a bidegree sequence* $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{N}_0^{(n,2)}$ *where* $\sum a_i = \sum b_i = n\bar{c}$ *and* $\min \mathbf{d} = m$*, where* $m \leq n-1$*. Let* $k_* = m+1+\sqrt{(m+1)^2 + n(\bar{c} - 2m)}$ *and define* $k = \lceil k_* \rceil$ *if* k_* *is real and* $k = 1$ *otherwise. If*

$$\max \mathbf{d} \leq \min(\left\lfloor n\frac{\bar{c} - m}{k} + m \right\rfloor, n-1),$$

then $\mathbf{d}$ *is graphic.*

Proof. Analogously to the proofs of Theorem 5 and Corollary 3, to construct the desired sufficient condition on M, we want the following inequalities to hold:

$$\sum_{i=1}^{n} a_i \leq (M+1)j \leq \sum_{i=1}^{n} \min(b_i, j).$$

Applying Lemma 1, it suffices to consider the case when $j = k$, where $\#(a_i = M) \leq k$. However, we know that for the remainder r in our usual minimizer construction, as given in equation (2.4), $r = n\bar{c} - Mk - m(n-k-1)$ or equivalently $kM + r - m + (n-k)m = n\bar{c}$, and consequently,

$$kM + k \leq n\bar{c} - (n-k)m + k.$$

Additionally we know that $nm + k(k-m) \leq \sum_i \min(b_i^*, k) \leq \sum_i \min(b_i, k)$, where $\mathbf{b}^*$ is the same as in the proof of Theorem 5. We construct the polynomials

$$R_*(k) = R(k) - k = k^2 - 2k(m+1) + 2nm - n\bar{c}$$

and

$$S_*(k) = S(k) - k - 1 = k^2 - 2m(k) + (2n-1)m - n\bar{c}.$$

Let k_* and k_{**} be the larger of the two roots of $R_*(k)$ and $S_*(k)$, respectively:

$$k_* = m + 1 + \sqrt{m^2 + 2m + 1 + n\bar{c} - 2nm},$$

$$k_{**} = m + \sqrt{m^2 + n\bar{c} + m - 2nm}.$$

It follows that if $k > k_*$, then both $R_*(k)$ and $S_*(k)$ are nonnegative. As before define $k = \lceil k_* \rceil$. Consequently, since $k(M - m) - nm \leq n\bar{c}$, we get the constraint that $M \leq \left\lfloor \frac{n\bar{c}-(n-k_*)m}{k} \right\rfloor$. This verifies the first k or, if there is a remainder, $k+1$ inequalities of the Gale-Ryser Theorem. As in the end of the proof of Theorem 5, invoking Lemma 1 will verify the remaining inequalities. □

Although the maximum value in Theorem 5 is easy to compute, it is not obvious if this bound is superior to both Theorem 3 (where $M_a = M_b$) and Theorem 2. Therefore we provide the following proof of superiority.

Corollary 4. *Consider degree sequences with a fixed minimum degree m, fixed average degree $\bar{c}$ such that $\bar{c} > m$, where we allow the number of nodes, n, in the sequence to vary. Notationally, for each $J \in \{2, 3, 4, 5, 6\}$, we can define $H_J(n, m, \bar{c})$ such that each Theorem J shows that a bidegree sequence is graphic (with loops) if the maximum degree $M \leq H_J(n, m, \bar{c})$. Then $\lim_{n\to\infty} \frac{H_q(n,m,\bar{c})}{H_p(n,m,\bar{c})} \geq 1$, for each $q \in \{5, 6\}$ and $p \in \{2, 3, 4\}$.*

Proof. We only prove the result for $H_5(n, m, \bar{c})$, although the proof for $H_6(n, m, \bar{c})$ is identical. We break the analysis up into two cases.

Case 1: $\bar{c} - 2m \leq 0$

In this case, $k \leq 2m$, and our condition on the maximum is $O(n)$, which is far superior to $O(\sqrt{n})$.

Case 2: $\bar{c} - 2m > 0$

Since we are only interested in asymptotic analysis, it suffices to consider the case when $k_* \in \mathbb{N}$ (so $k = k_*$) and $H_5(n, m, \bar{c}) = n\frac{\bar{c}-m}{k} + m \in \mathbb{N}$. Consequently,

$$H_5(n, m, \bar{c}) = n(\sqrt{m^2 + n(\bar{c} - 2m)} - m)\frac{\bar{c} - m}{n(\bar{c} - 2m)} + m$$

$$= (\sqrt{m^2 + n(\bar{c} - 2m)} - m)\frac{\bar{c} - m}{(\bar{c} - 2m)} + m$$

$$= \sqrt{n\frac{(\bar{c}-m)^2}{\bar{c}-2m} + m^2(\frac{\bar{c}-m}{\bar{c}-2m})^2} - m(\frac{m}{\bar{c}-2m}).$$

Note that asymptotically for large n, fixed m and $\bar{c}$, if $p = 2$, then

$$\lim_{n\to\infty} \frac{H_p}{\sqrt{4mn}} = 1,$$

while if $p \in \{3, 4\}$, then

$$\lim_{n\to\infty} \frac{H_p}{\sqrt{\bar{c}n}} \leq 1.$$

So asymptotically, to demonstrate that Theorem 5 is indeed more powerful than Theorems 2-4, we want to show that $\frac{(\bar{c}-m)^2}{\bar{c}-2m} \geq \bar{c}$ and $\frac{(\bar{c}-m)^2}{\bar{c}-2m} \geq 4m$. For the ensuing discussion, let $\bar{c} = x$, $m = y$, with $x > 2y$ by assumption.

First consider $\frac{(x-y)^2}{x-2y} \geq x \iff x^2 - 2xy + y^2 \geq x^2 - 2yx$, which is true always. Next, note that $\frac{(x-y)^2}{x-2y} \geq 4y \iff x^2 - 2xy + y^2 \geq 4xy - 8y^2 \iff x^2 \geq 6xy - 9y^2$. Since $x > 0$, this inequality is equivalent to $1 \geq 6(\frac{y}{x}) - 9(\frac{y}{x})^2$. Using another change of variables, where $a = \frac{y}{x}$, we want to know when $1 \geq 6a - 9a^2$. Taking the derivative of the right hand side implies that the maximum value of the right hand side occurs at $a = \frac{1}{3}$. Since $6(\frac{1}{3}) - 9(\frac{1}{9}) = 1$, we conclude that asymptotically, Theorem 5 is more powerful than Theorems 2-4. □

As a simple example, note that if $m = 1$, $\bar{c} = 4$, $n = 10$, and $M = 6$ then $k = 6$ and $M \leq \lfloor n(\bar{c} - m)/k + m \rfloor = 6$, so Theorem 5 holds, but $(m + M)^2/4 = \frac{49}{4} > 12 > 10 = mn$ so Theorem 2 fails.

As a final comment regarding Theorem 5, while it is not surprising that we can sharpen the bounds on the maximum by including an additional parameter ($n\bar{c}$), corresponding to the total number of edges, it is not readily apparent why the bound would dramatically change from $O(\sqrt{n})$ to $O(n)$ as *two* times the minimum number of edges of a node approaches the average number of edges.

We now conclude our results section with a corollary of Theorem 5 that yields a more flexible graphicality criterion in which the degrees of some nodes can exceed the upper bound mentioned in Theorem 5.

Corollary 5. *Consider a bidegree sequence* $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{N}_0^{(n,2)}$ *where* $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i = n\bar{c}$ *and* $\min \mathbf{d} = m$*, with* $m \leq n$ *and* $\max \mathbf{d} \leq n$*. Without loss of generality, take the* a_i *to be arranged in non-increasing order. Assume that there exists an* R *such that* $\sum_{i=1}^R a_i = n\lambda$*, and* $\sum_{i=1}^R b_i \leq n\lambda$ *where* $\lambda < m$ *and* $n - n\frac{\lambda}{m} - R \geq 1$*. Next, define* $M = \max_{i \geq R} \max(a_i, b_i)$ *and* $k_* = m + \sqrt{m^2 + n(\bar{c} - 2m) + Rm}$*. Let* $k = \lceil k_* \rceil$ *if* k_* *is real and* $k = 1$ *otherwise. If* $M \leq \min(\left\lfloor \frac{n\bar{c} - nm - n\lambda + Rm}{k} + m \right\rfloor, n)$ *and if either* $k \leq M$ *or* $k \leq n - n\frac{\lambda}{m} - R$*, then* $\mathbf{d}$ *is graphic with loops.*

Proof. The proof is quite similar to that of Theorem 5, so we only provide a sketch and leave the details to the reader. Given that we defined $\sum_{i=1}^R a_i = n\lambda$, and $\lambda < m$, the first R inequalities of the Gale-Ryser Theorem are trivially satisfied. Furthermore, the first m inequalities are satisfied trivially as well since $\max \mathbf{d} \leq n$.

As in the proof of Theorem 5, we note that for arbitrary $k > 0$, $\sum_{i=1}^{k+R} a_i \leq kM + n\lambda$. For our minimizing degree sequence, $n\bar{c} = kM + (r - k) + (n - k)m + n\lambda - Rm$, where r is defined in the proof of Theorem 5 and hence $kM \leq n\bar{c} - (n - k)m + Rm - n\lambda$ since $r \geq m$. Thus, $kM + n\lambda \leq n\bar{c} + Rm - (n - k)m$.

Similarly, since we can assume that $k > m$, it follows that
$nm + k(k - m) \leq \sum_{i=1}^n \min(b_i, k) \leq \sum_{i=1}^n \min(b_i, k + R)$, provided that $k \leq M$.

Putting the bounds on $\sum_{i=1}^{k+R} a_i$ and $\sum_{i=1}^n \min(b_i, k+R)$ together, to satisfy the remainder of the first $k + R$ Gale-Ryser inequalities, under the assumption that $k \leq M$, we want to fulfill the inequality $nm + k(k - m) - n\bar{c} - Rm + (n - k)m \geq 0$, where equality is achieved when $k = m + \sqrt{m^2 + n(\bar{c} - 2m) + Rm}$. Consequently, if $M \leq \left\lfloor \frac{n\bar{c} - n\lambda - (n-R)m}{\lceil k \rceil} + m \right\rfloor$, then the first $k + R \leq M$ inequalities in the Gale-Ryser Theorem will be satisfied. To finish off the proof, we then consider the case where $k > M$. We know that $\sum_{i=1}^n \min(b_i, k + R) \geq n\bar{c} - n\lambda$ and $\sum_{i=1}^{k+R} a_i \leq n\bar{c} - (n - k - R)m$. Consequently, we require that $n\lambda + Rm \leq (n - k)m$, or equivalently, $k \leq (n - n\frac{\lambda}{m} - R)$. Hence, our assumptions imply that the first M inequalities hold.

Now suppose for simplicity that $r = 0$. For the degree sequence that maximizes the in-degree vector $\mathbf{a}$ in the Gale-Ryser Theorem, $a_j = m$ for all $j > k + R$, and hence $\sum_{i=1}^j a_i$ grows linearly in j for these j. We can therefore complete the proof by invoking Lemma

1, since $\sum_{i=1}^{n} \min(b_i, k)$ is concave in k. This result implies that the remaining inequalities must hold. In the case where $r > 0$, as before in Theorems 5 and 6, we exploit the existence of the remainder to construct refined inequalities that demonstrate that our prior choice for M is indeed correct. $\square$

Recalling Counterexample 1, the only way we were able to construct a degree sequence that was not graphic was by having many nodes with degrees greater than $\sqrt{n\bar{c}}$. In contrast, Corollary 5 tells us that in an asymptotic sense, as long as we have a relatively small number of nodes R with degrees that surpass $O(\sqrt{n})$, such that the sum of their degrees is $n\lambda = O(n^{1-\tau})$ for some $\tau > 0$, then asymptotically we still have graphicality provided that $O(n)$ nodes are bounded in-degree by essentially the same bound derived in Theorems 5 and 6. This observation is useful, for example, for broadening the graphicality criteria for so-called scale free networks with exponent greater than 2. For such networks, we find that the expected number of edges contributed by nodes of degree greater than $\sqrt{n}$ is $n \int_{\sqrt{n}}^{n} \frac{x}{x^{2+\tau}} dx = O(n^{1-\frac{\tau}{2}})$. In this setting, Corollary 5 can be viewed in parallel with the prior work of Chen and Olvera-Cravioto [23], who proved that provided the sum of the in-degrees equals the sum of the out-degrees, randomly generated degree sequences from a scale-free distribution with a finite mean (that is, with an exponent greater than 2) are asymptotically (almost surely) graphic.

2.3 DISCUSSION

While the famous Gale-Ryser inequalities (e.g., [10, 57]) provide necessary and sufficient conditions for a degree sequence to be graphic, checking these inequalities and using them to generate graphs [49] can be computationally inefficient. Work by Zverovich, Alon, Cairns and their collaborators provides simplified sufficient conditions for graphicality; however, these conditions assume that the in-degree vector equals the out-degree vector for directed graphs and are posed in terms of the minimum and maximum of the degree sequence. In our analysis, we drop the assumption that the in-degree vector equals the out-degree vector and prove an alternative sufficient condition for graphicality incorporating the average degree

(Theorems 3, 5). We prove that for fixed minimum and average degree, for sufficiently large n, Theorem 5 provides more flexible conditions to demonstrate graphicality than those provided by prior work. The proof method used in this paper builds heavily on that used by Dahl and Flatberg [31] and Miller [57] in their approaches to relaxing the graphicality conditions in the Erdös-Gallai and Gale-Ryser Theorems, with the key idea being to exploit the discrete concavity of the functions appearing in the relevant inequalities. Note that while all results in this paper are stated in terms of bidegree sequences for directed graphs, these results apply immediately to bipartite graphs, while the proof methods will extend directly to the case of undirected graphs as well.

In Counterexample 1, we show that we cannot expect to do much better than our sufficient conditions for graphicality using bounds on the average degree alone. However, we also notice that to construct a degree sequence that is not graphic, we must choose many nodes to have large degree. This observation motivates Corollary 5, which says that as long as only a relatively small number of node degrees exceed $O(\sqrt{n})$, we still have graphicality. Interpreted in an asymptotic sense, we can relate this result to the work of Chen and Olvera-Cravioto [23], which shows that asymptotically, degree sequences generated from scale-free distributions with exponent greater than 2 almost surely will be graphic.

3.0 ASYMPTOTIC ENUMERATION OF GRAPHS WITH PRESCRIBED DEGREE SEQUENCES

Given a degree sequence, a fundamental question to ask is whether it is graphic; that is, does there exist a graph for which the degrees of the nodes are exactly the elements of the sequence? A more refined view goes beyond simply considering graphicality as a yes-or-no property and recognizes that there may be very different numbers of graphs that realize different graphic degree sequences. Our main goal in this work is to develop formulas that approximate the numbers of graphs that realize degree sequences with certain properties, which are valid asymptotically as the number of nodes in the degree sequence and in the corresponding graphs goes to infinity.

The problem of counting graphs that realize a given degree sequence can be recast as a problem of counting $0-1$ binary matrices. Specifically, counting the number of rectangular $0-1$ binary matrices with fixed row and column sums is equivalent to counting the number of bipartite graphs with a fixed bidegree sequence. Alternatively, counting the number of square $0-1$ binary matrices with fixed row and column sums is equivalent to counting the number of directed graphs with loops that realize a given bidegree sequence. Since any rectangular $0-1$ binary matrix can be arbitrarily extended to a square $0-1$ binary matrix by adding either rows or columns of $0's$, we focus here on square binary matrices.

In terms of applications, the aforementioned counting problem is an important step for uniformly generating $0-1$ matrices (contingency tables) with fixed row and column sums. Uniform generation of $0-1$ binary matrices has many applications, from detecting statistically significant subgraphs (motifs) in a network in data mining [46, 41, 26] to determining the impact of the degree sequence on emergent dynamics in a network of nodes with temporally varying states [80].

We provide a novel method for constructing asymptotics (to arbitrary accuracy) for the number of directed graphs that realize a fixed bi-degree sequence $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{Z}^{N\times 2}$ with maximum degree $d_{max} = O(S^{\frac{1}{2}-\tau})$ for an arbitrarily small positive number τ, where S is the number edges specified by $\mathbf{d}$. Our approach is based on two key steps, graph partitioning and degree preserving switches. The former idea allows us to relate enumeration results for given sequences to those for sequences that are especially easy to handle, while the latter facilitates expansions based on numbers of shared neighbors of pairs of nodes. While we focus primarily on directed graphs allowing loops, our results can be extended to other cases, including bipartite graphs, as well as directed and undirected graphs without loops. In addition, we can relax the constraint that $d_{max} = O(S^{\frac{1}{2}-\tau})$ and replace it with $a_{max}b_{max} = O(S^{1-\tau})$, where a_{max} and b_{max} are the maximum values for $\mathbf{a}$ and $\mathbf{b}$ respectively. The previous best results, from Greenhill et al. [42], only allow for $d_{max} = o(S^{\frac{1}{3}})$ or alternatively $a_{max}b_{max} = o(S^{\frac{2}{3}})$. Since in many real world networks, d_{max} scales larger than $o(S^{\frac{1}{3}})$, we expect that this work will be helpful for various applications.

To attain this generality, we initially estimate the ratio $\|G_{\mathbf{d_1}}\|/\|G_{\mathbf{d_2}}\|$, where $\|G_{\mathbf{d_i}}\|$ is the number of directed graphs with loops that realize the bidegree sequence $\mathbf{d}_i$, $i = 1, 2$, under the constraint that the maximum degree for both of these bidegree sequences is $O(S^{\frac{1}{2}-\tau})$ for any $\tau \in (0, \frac{1}{2}]$. We can estimate this ratio accurately when the taxicab norm of the difference of the bidegree sequences, $\|\mathbf{d_1} - \mathbf{d_2}\|_1$, equals 2, and this relation is assumed in the statements of the theorems that we prove. We can apply the theorems in this work to estimate $\|G_{\mathbf{d_{k+1}}}\|/\|G_{\mathbf{d_0}}\|$, where $\|\mathbf{d_{k+1}} - \mathbf{d_0}\|_1 > 2$, by considering a product $\Pi_{i=0}^{k}\|G_{\mathbf{d_{i+1}}}\|/\|G_{\mathbf{d_i}}\| = \|G_{\mathbf{d_{k+1}}}\|/\|G_{\mathbf{d_0}}\|$, where for all i, $\|\mathbf{d_i} - \mathbf{d_{i+1}}\|_1 = 2$. (For a rigorous proof that we can construct a sequence of $\mathbf{d_i}$'s in this fashion, we refer the reader to a result by Muirhead that can be found in [11].)

We now summarize our main results as follows:

- In Theorem 7, we derive an expansion for $\|G_{\mathbf{d}}\|$ that holds in general for all degree sequences.
- In Corollary 8, we exploit the sparsity constraints to prove that the terms in the expansion of $\|G_{\mathbf{d_1}}\|/\|G_{\mathbf{d_2}}\|$ based on Theorem 7 decrease geometrically.
- Starting with Corollary 9, we establish an asymptotic approximation for $\|G_{\mathbf{d_1}}\|/\|G_{\mathbf{d_2}}\|$

allowing for errors of size $O(S^{-2\tau})$.

- Then under modest assumptions regarding the asymptotic approximation for $\|G_{\mathbf{d_1}}\|/\|G_{\mathbf{d_2}}\|$ where we allow errors of $O(S^{-2w\tau})$, for some positive integer w, in Theorem 9 we provide a general method that yields an approximation for $\|G_{\mathbf{d_1}}\|/\|G_{\mathbf{d_2}}\|$ allowing for errors of size $O(S^{-2\tau-2w\tau})$.
- Next in Theorem 10, we demonstrate that if we know that our approximation for $\|G_{\mathbf{d_1}}\|/\|G_{\mathbf{d_2}}\|$ only allow errors of $O(S^{-\gamma})$ where $\frac{1}{2} \leq \gamma$, we do not need the 'modest assumptions' made in Theorem 9 to derive a sharper approximation of $\|G_{\mathbf{d_1}}\|/\|G_{\mathbf{d_2}}\|$ where our new error term is now $O(S^{-\gamma-2\tau})$. (As the proofs of Theorems 9 and 10 are very similar, we place the proof in Section 3.6.)
- Then using Theorem 11 we demonstrate how we can recover an (arbitrarily) accurate asymptotic approximation of $\|G_{\mathbf{d_1}}\|$ with knowledge of an (arbitrarily) accurate approximation of the ratio, $\|G_{\mathbf{d_1}}\|/\|G_{\mathbf{d_2}}\|$
- Subsequently, in Section 3.4 we show that the 'modest assumptions' of Theorem 9 hold and use that result to establish successively finer asymptotic approximations for $\|G_{\mathbf{d_1}}\|/\|G_{\mathbf{d_2}}\|$, allowing errors of size $O(S^{-4\tau})$,$O(S^{\max(-\frac{1}{2}-3\tau,-6\tau)})$, and $O(S^{\max(-1-2\tau,-8\tau)})$ respectively, where S is the number of edges in the graph and $\tau \in (0, \frac{1}{2}]$.
- Finally, in Section 3.5 we demonstrate that the 'modest assumptions' in Theorem 9 do in fact hold if our approximation allows for errors of size $O(S^{-\frac{1}{2}})$. This result combined with Theorems 9 and 10 prove that our method enables us to derive approximations of $\|G_{\mathbf{d_1}}\|/\|G_{\mathbf{d_2}}\|$ up to arbitrary accuracy.
- In Section 3.7, we explain how to generalize our results to the case where the product of the maximum in-degree and maximum out-degree is of $O(S^{1-\tau})$. and illustrate how to extend our results to undirected and directed graphs, including the case where loops are prohibited, as well as graphs where we prohibit edges between certain nodes. We also explore how to compute the likelihood that two arbitrary nodes share an edge with each other.

We now detail our proof strategy. First, in Section 3.1, we use *graph partitioning*, inspired by [58], to construct novel expansions for the number of graphs that realize a bidegree sequence. That is, for a particular realization of a bidegree sequence, we can partition our adjacency matrix into two submatrices, one submatrix containing just the ith and jth rows (or columns), and another submatrix containing the remaining $N-2$ rows. In turn, we obtain two “smaller” bidegree sequences for both of our submatrices. Once we demonstrate how to count the number of graphs that realize the smaller bidegree sequence corresponding to the two-row submatrix, we obtain the following expression for $\|G_{\mathbf{d}}\|$ in terms of the number of graphs where two arbitrarily chosen nodes i and j (with degrees a_i,a_j) have k common neighbors:

$$\|G_{\mathbf{d}}\| = \sum_{k=0}^{a_j} \{ \binom{a_i + a_j - 2k}{a_j - k} \sum_{\mathbf{r} \in X_k} \|G_{\mathbf{r}}\| \} \tag{3.1}$$

where $\mathbf{r}$ is a (residual) bidegree sequence of the $N-2$ remaining rows and X_k is a set of (residual) bidegree sequences corresponding to graphs where the ith and jth nodes have exactly k common neighbors.

Next, in Section 3.2, we introduce the idea of *degree preserving switches*, as discussed in [56, 55, 59, 69], in which we make a single edge replacement to eliminate a common neighbor of two nodes without changing any nodes' degrees. Counting graphs with common images or pre-images under degree preserving switches allows us to prove that for sparse graphs, the dominant term in the expansion only involves instances where there are no common neighbors between the two nodes i and j. That is, in the notation of (3.1),

$$\|G_{\mathbf{d}}\| \approx \binom{a_i + a_j}{a_j} \sum_{\mathbf{r} \in X_0} \|G_{\mathbf{r}}\|.$$

Moreover, it turns out that the set X_0 in the prior expression does not change if we consider $\mathbf{d}_*$ where $\mathbf{d}_* = \mathbf{d}$ except that node i in $\mathbf{d}_*$ has degree $a_i - 1$ instead of a_i for $\mathbf{d} = (\mathbf{a}, \mathbf{b})$ with $\mathbf{a} = (a_1, \ldots, a_N)^T$. Consequently, we find that

$$\frac{\|G_{\mathbf{d}_*}\|}{\|G_{\mathbf{d}}\|} \approx \binom{a_i + a_j - 1}{a_i - 1} \Big/ \binom{a_i + a_j - 1}{a_j - 1} = \frac{a_i}{a_j}.$$

A subtlety of the proof is that we first establish the above statement where node i or j has bounded (in)-degree. We then show that the above relationship still holds even for degree sequences that lack a node of bounded degree by using the fact that such degree sequences are very close in taxicab norm to a degree sequence that contains a node of bounded degree.

We then proceed in Section 3.3 by proving (under modest assumptions) a general technique that allows for more refined approximations of $\|G_d\|$, both by considering instances where nodes i and j can have a nonzero number of common neighbors. We also show how asymptotics for $\|G_{\mathbf{d}}\|$ follow from those obtained for the ratio $\|G_{\mathbf{d}}\|/\|G_{\mathbf{d}_*}\|$, where $\mathbf{d}_*$ is a degree sequence designed such that both this ratio and $\|G_{\mathbf{d}_*}\|$ itself can be estimated. More precisely, we achieve this result by working with a sequence of intermediary degree sequences, starting from a sequence for which it is easy to compute the number of graphs that realize it and, from there, progressing successively closer to $\mathbf{d}$

Subsequently in Section 3.4, we put the technique proposed by Section 3.3 into practice and computationally verify that the 'modest assumptions' do indeed hold and derive various asymptotic approximations for the ratio $\|G_{\mathbf{d}}\|/\|G_{\mathbf{d}_*}\|$. And finally in Section 3.5, we prove that these 'modest assumptions' hold and consequently that our method yields arbitrarily accurate asymptotic approxiations for the ratio $\|G_{\mathbf{d}}\|/\|G_{\mathbf{d}_*}\|$.

3.1 COUNTING GRAPHS WITH PARTITIONING

In this paper, we will consider *bidegree sequences*, each denoted by $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{Z}^{N\times 2}$, where for concreteness we specify that $\mathbf{a}$ lists the in-degrees of the nodes in the sequence and $\mathbf{b}$ the out-degrees. We are interested in graphs that realize such sequences, where we allow either 0 or 1 connection between each pair of nodes as well as single self-loops within each graph. Throughout the paper, we use S to denote the number of edges in the graphs that realize $\mathbf{d}$. Since we do not need to distinguish between degree sequences and bidegree sequences in this work, we will henceforth simply refer to these as degree sequences. In one special case, it is particularly easy to count the number of graphs that realize a specified degree sequence.

Lemma 2. *Suppose that*

$$\mathbf{a} = \{a_1, \ldots, a_k, 0, \ldots, 0\}, \mathbf{b} = \{k, .., k, 1, .., 1, 0, .., 0\}, \quad \textit{and} \quad \sum_{i=1}^{k} a_i = \sum_{i=1}^{N} b_i =: S.$$

Let q denote the number of times k appears in $\mathbf{b}$ and let $p = \min\{a_1, \ldots, a_k\}$. If $p < q$, then there does not exist a graph that realizes this bidegree sequence. If $p \geq q$ then there are

$$\frac{(S-qk)!}{\Pi_i(a_i-q)!}$$

graphs that realize the bidegree sequence. Similarly, if $\mathbf{a} = \{k, .., k, 1, .., 1, 0, .., 0\}$ and $\mathbf{b} = \{b_1, \ldots, b_k, 0, \ldots, 0\}$, with corresponding definitions of p and q, then there are

$$\frac{(S-qk)!}{\Pi_i(b_i-q)!}$$

graphs that realize the bidegree sequence.

Proof. We present the proof for the first case, since the second is completely analogous. We first note that the q nodes in $\mathbf{b}$ with out-degrees equal to k must connect to all of the nodes with nonzero degree in $\mathbf{a}$. Of the $S - qk$ remaining outward edges, start with the node that corresponds to a_1. There are $S - qk$ choices for outward edges that can supply the $a_1 - q$ unconnected inward edges to this node, such that there are $\binom{S-qk}{a_1-q}$ possible ways to link this node into the graph. Once these $a_1 - q$ edges have been connected, there are $\binom{S-qk-a_1+q}{a_2-q}$ ways to link the node that corresponds to a_2 into the graph. Notice that

$$\binom{S-qk-a_1+q}{a_2-q}\binom{S-qk}{a_1-q} = \frac{(S-qk)!}{(a_1-q)!(a_2-q)!(S-qk-a_1-a_2+2q)!}.$$

Multiplying inductively, the $(S - qk - a_1 - a_2 + 2q)!$ term disappears and we obtain the desired result. □

Remark 1. *Lemma 2 generalizes immediately to arbitrary permutations of the given $\mathbf{a}$ and $\mathbf{b}$.*

With the above lemma in hand, we can construct an inductive characterization for the number of graphs that realize a specified degree sequence. First, though, we must define some notation.

Definition 2. *Let $\|G_{\mathbf{d}}\|$ denote the number of graphs that realize a degree sequence $\mathbf{d}$. Furthermore, for a set X of degree sequences, let $\|G_X\|$ be the number of graphs that realize any degree sequence in X.*

Now, consider an arbitrary adjacency matrix $\mathbf{A} \in \{0,1\}^{M \times N}$. We can write $\mathbf{A}$ as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_N \\ \delta_{M-1,1} & \delta_{M-1,2} & \cdots & \delta_{M-1,N} \\ \delta_{M,1} & \delta_{M,2} & \cdots & \delta_{M,N} \end{bmatrix}$$

where for each i, $\mathbf{A}_i \in \{0,1\}^{(M-2)\times 1}$. Of course, letting $\mathbf{0}$ denote a column vector of $M-2$ zeros, we have

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_N \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \delta_{M-1,1} & \delta_{M-1,2} & \cdots & \delta_{M-1,N} \\ \delta_{M,1} & \delta_{M,2} & \cdots & \delta_{M,N} \end{bmatrix}.$$

Call the first and second matrices in this equation $\mathbf{A}_l$ and $\mathbf{A}_r$, respectively. Now, if $\mathbf{A}$ realizes a given degree sequence, $(\mathbf{a}, \mathbf{b})$, then there is a vector $\{s_1, ..., s_N\}$ such that the degree sequence of (the graph corresponding to) $\mathbf{A}_r$ is $(\{0, ..., 0, a_{M-1}, a_M\}, \{s_1, ..., s_N\})$ and the degree sequence of $\mathbf{A}_l$ is $(\{a_1, ..., a_{M-2}, 0, 0\}, \{b_1 - s_1, ..., b_N - s_N\})$, with the constraint that none of the s_i (i.e., the column sums of $\mathbf{A}_r$) can exceed 2 and the s_i must sum to $a_{M-1} + a_M$. If we know the number of s_i that equal 2, we can invoke Lemma 1 to count the number of realizations of the degree sequence of $(\{0, ..., 0, a_{M-1}, a_M\}, \{s_1, ..., s_N\})$. This idea, extended to a partition of the ith and jth rows rather than specifically the $(M-1)$st and Mth rows, motivates a useful theorem. To state the theorem, we define the set of degree sequences

$$\begin{aligned} X_k(i, j; \mathbf{d}) = \{(\mathbf{a} - a_i\mathbf{e_i} - a_j\mathbf{e_j}, \mathbf{b} - \mathbf{s}) : \; & \#(s_n = 2) = k, \; \#(s_n \geq 3) = 0, \\ & \text{and } \textstyle\sum_{n=1}^{N} s_n = a_i + a_j\}. \end{aligned} \tag{3.2}$$

Note that in equation (3.2), we explicitly represent the positions (i, j) from which edges will be partitioned out as well as the base degree sequence $\mathbf{d} = (\mathbf{a}, \mathbf{b})$. $X_k(i, j; \mathbf{d})$ is a set of degree sequences, even with $\mathbf{d}$ fixed, because different choices of $\mathbf{s}$ can be made that fit the definition in (3.2). The notation $X_k(i, j; \mathbf{d})$ is cumbersome and the arguments of X_k will be dropped when possible, but this notation will be needed to make results precise later in the paper.

Theorem 7. *Fix a degree sequence* $\mathbf{d} = (\mathbf{a}, \mathbf{b})$*. Pick an arbitrary pair of nodes, say with indices* i, j *and corresponding in-degrees* a_i, a_j*, where* $a_j \le a_i$*, and define* $X_k = X_k(i, j; \mathbf{d})$ *as in (3.2). Then*

$$\|G_{\mathbf{d}}\| = \sum_{k=0}^{a_j} \binom{a_i + a_j - 2k}{a_j - k} \|G_{X_k}\|.$$

Proof. Any adjacency matrix that realizes $\mathbf{d}$ can be partitioned into two adjacency matrices as we have discussed. Picking any two nodes, with in-degrees denoted by a_i, a_j, we note that any realization of our degree sequence $\mathbf{d}$ must also be a realization of some degree sequence $(\mathbf{a} - a_i\mathbf{e_i} - a_j\mathbf{e_j}, \mathbf{b} - \mathbf{s}) \in X_k$ for some $k \le N$, combined with a realization of the degree sequence $(a_i\mathbf{e_i} + a_j\mathbf{e_j}, \mathbf{s})$. In order for $(a_i\mathbf{e_i} + a_j\mathbf{e_j}, \mathbf{s})$ to be graphic, we require that $\#(s_i = 2) \le a_j$ (see Lemma 1), hence only graphs that realize degree sequences in X_k for $k \le a_j$ can correspond to our adjacency matrix partition. If $(\mathbf{a} - a_i\mathbf{e_i} - a_j\mathbf{e_j}, \mathbf{b} - \mathbf{s}) \in X_k$ for $k \le a_j$, then Lemma 1 implies that the number of graphs that realize $(a_i\mathbf{e_i} + a_j\mathbf{e_j}, \mathbf{s})$ is precisely $\binom{a_i+a_j-2k}{a_j-k}$. By multiplying this quantity by the number of graphs that can be generated by the residual degree sequence $(\mathbf{a} - a_i\mathbf{e_i} - a_j\mathbf{e_j}, \mathbf{b} - \mathbf{s})$, namely $\|G_{X_k}\|$, and summing over all X_k for $k \le a_j$, we obtain the desired result. □

Remark 2. *Based on Definition 1,* $\|G_{X_k}\|$ *represents the expression* $\sum_{\mathbf{r} \in X_k} \|G_{\mathbf{r}}\|$*, such that Theorem 7 agrees with equation (3.1).*

At this point, we introduce some additional notation. The rationale is that we will want to compare numbers of graphs realizing two different degree sequences that are identical except that the in-degrees of two particular nodes in the sequences differ by 1. A succint way to think of this relation is to start with a degree sequence $\mathbf{d} = (\mathbf{a}, \mathbf{b})$ for which $\sum_k a_k = \sum_k b_k + 1$. Now, for any entry $a_k \ge 1$ in $\mathbf{a}$, let $\mathbf{a_{-k}} = \mathbf{a} - \mathbf{e_k}$, $\mathbf{d_{-i}} = (\mathbf{a_{-i}}, \mathbf{b})$, and $\mathbf{d_{-j}} = (\mathbf{a_{-j}}, \mathbf{b})$. Note that the sums of the in-degrees for the sequences $\mathbf{d_{-i}}$ and $\mathbf{d_{-j}}$ are identical and are equal to the sum of their out-degrees, and $\mathbf{d_{-i}}, \mathbf{d_{-j}}$ are related as desired. Given Theorem 7, we can attain a nontrivial approximation for the ratio of the number of graphs that realize the degree sequence $\mathbf{d_{-i}}$ compared to the number of graphs that realize the degree sequence $\mathbf{d_{-j}}$. To do so, we will need to work with $X_k(i, j; \mathbf{d_{-i}})$ and $X_k(i, j; \mathbf{d_{-j}})$. But note that since $\mathbf{d_{-i}}, \mathbf{d_{-j}}$ are both defined from the same base sequence $\mathbf{d}$ and have the same in-degree sum, these two X_k are identical; hence, we will simply refer to this set as X_k in the proof below, and we will similarly drop the arguments of X_k in most subsequent parts of the paper.

Corollary 6. *If $a_i \geq a_j > 0$, then*

$$\|G_{\mathbf{d_{-i}}}\| \geq \frac{a_i}{a_j}\|G_{\mathbf{d_{-j}}}\|.$$

Proof. First define the variable δ such that if $a_j + 1 \leq a_i$, then $\delta = 1$ and otherwise, $\delta = 0$. The statement of the Corollary holds trivially if $\mathbf{d_{-j}}$ is not graphic. If $\mathbf{d_{-j}}$ is graphic, then so is $\mathbf{d_{-i}}$ and we can apply Theorem 1 and note that

$$\|G_{\mathbf{d_{-i}}}\| = \sum_{k=0}^{a_j-1+\delta} \binom{(a_i-1)+a_j-2k}{a_j-1+\delta-k}\|G_{X_k}\| \geq \sum_{k=0}^{a_j-1} \binom{a_i+a_j-2k-1}{a_j-k}\|G_{X_k}\|.$$

Similarly, $\|G_{\mathbf{d_{-j}}}\| = \sum_{k=0}^{a_j-1}\binom{a_i+a_j-2k-1}{a_j-k-1}\|G_{X_k}\|$.

If we show that for all relevant natural numbers k,

$$\binom{a_i+a_j-2k-1}{a_j-k} \geq \frac{a_i}{a_j}\binom{a_i+a_j-2k-1}{a_j-k-1},$$

then the result will follow. Note that

$$\binom{a_i+a_j-2k-1}{a_j-k} \Big/ \binom{a_i+a_j-2k-1}{a_j-k-1} = \frac{a_i-k}{a_j-k} \geq \frac{a_i}{a_j},$$

since $a_i \geq a_j$, and the proof is complete.

□

We next seek a compact expression for $\|G_{\mathbf{d_{-i}}}\|/\|G_{\mathbf{d_{-j}}}\|$ that will enable us to analyze this ratio with minimal difficulty without having to explicitly worry about the number of terms in the formulas for $\|G_{\mathbf{d_{-i}}}\|$ or $\|G_{\mathbf{d_{-j}}}\|$ found in the summation in Theorem 1. Before we do that we introduce two additional notational conventions:

$$\Pi_{k=r}^{r-1}\omega_k = 1 \quad \text{and} \quad \Pi_{k=0}^{0}\omega_k = \omega_0. \tag{3.3}$$

Corollary 7.

$$\frac{\|G_{\mathbf{d_{-i}}}\|}{\|G_{\mathbf{d_{-j}}}\|} = \frac{a_i}{a_j}\left(\frac{\sum_{k=0}^{a_j}\Pi_{l=0}^{k-1}(a_j-l)\Pi_{l=1}^{k}(a_i-l)\eta_k}{\sum_{k=0}^{a_j-1}\Pi_{l=1}^{k}(a_j-l)\Pi_{l=0}^{k-1}(a_i-l)\eta_k}\right) \tag{3.4}$$

where

$$\eta_k = \frac{\|G_{X_k}\|}{[\Pi_{l=0}^{2k-1}(a_i+a_j-l-1)]\|G_{X_0}\|} \tag{3.5}$$

when $k \leq \lfloor\frac{a_i+a_j-1}{2}\rfloor$, *and* $\eta_{a_j} \equiv \eta_{a_i} = 1$ *when* $k = a_j$ *if* $a_j = a_i$.

Proof. Without loss of generality, assume $a_j < a_i - 1$. That is, if $a_j \in \{a_i, a_i - 1\}$, then we can adjust the calculations with $\delta = 1$ as in the proof of Corollary 1 and they will go through; furthermore, if $a_j = a_i$, then the $k = a_j = a_i$ term in the numerator of the right hand side of (3.4) is 0, so we can set $\eta_{a_i=a_j} = 1$ (or any finite number) and the result will still hold.

By invoking Theorem 1, using $X_k = X_k(i,j;\mathbf{d_{-i}}) = X_k(i,j;\mathbf{d_{-j}})$, and dividing the numerator and denominator by $\|G_{X_0}\|$, we obtain

$$\begin{aligned}&\|G_{\mathbf{d_{-i}}}\|/\|G_{\mathbf{d_{-j}}}\| = \\ &\left(\sum_{k=0}^{a_j}\binom{a_i+a_j-2k-1}{a_j-k}\|G_{X_k}\|/\|G_{X_0}\|\right) / \left(\sum_{k=0}^{a_j-1}\binom{a_i+a_j-2k-1}{a_j-k-1}\|G_{X_k}\|/\|G_{X_0}\|\right).\end{aligned} \tag{3.6}$$

Now we multiply both the numerator and denominator by $\frac{a_i!a_j!}{(a_i+a_j-1)!}$, which yields

$$\begin{aligned}&\|G_{\mathbf{d_{-i}}}\|/\|G_{\mathbf{d_{-j}}}\| = \\ &\left(\sum_{k=0}^{a_j}\frac{\Pi_{l=0}^{k-1}(a_j-l)\Pi_{l=0}^{k}(a_i-l)\|G_{X_k}\|}{\Pi_{l=0}^{2k-1}(a_i+a_j-l-1)\|G_{X_0}\|}\right) / \left(\sum_{k=0}^{a_j-1}\frac{\Pi_{l=0}^{k}(a_j-l)\Pi_{l=0}^{k-1}(a_i-l)\|G_{X_k}\|}{\Pi_{l=0}^{2k-1}(a_i+a_j-l-1)\|G_{X_0}\|}\right).\end{aligned} \tag{3.7}$$

Substituting in η_k as defined in equation (3.5), we obtain

$$\begin{aligned}&\|G_{\mathbf{d_{-i}}}\|/\|G_{\mathbf{d_{-j}}}\| = \\ &\frac{\sum_{k=0}^{a_j}\Pi_{l=0}^{k-1}(a_j-l)\Pi_{l=0}^{k}(a_i-l)\eta_k}{\sum_{k=0}^{a_j-1}\Pi_{l=0}^{k}(a_j-l)\Pi_{l=0}^{k-1}(a_i-l)\eta_k} = \frac{a_i}{a_j}\left(\frac{\sum_{k=0}^{a_j}\Pi_{l=0}^{k-1}(a_j-l)\Pi_{l=1}^{k}(a_i-l)\eta_k}{\sum_{k=0}^{a_j-1}\Pi_{l=1}^{k}(a_j-l)\Pi_{l=0}^{k-1}(a_i-l)\eta_k}\right),\end{aligned} \tag{3.8}$$

which agrees with equation (3.4). □

Remark 3. *Obviously the right hand side of (3.4) does not change if we add zero terms to the sums in the numerator and denominator. The summation in the numerator of this expression can be rewritten using*

$$\sum_{k=0}^{a_j} \Pi_{l=0}^{k-1}(a_j - l)\Pi_{l=1}^{k}(a_i - l)\eta_k = \sum_{k=0}^{a_j+1} \Pi_{l=0}^{k-1}(a_j - l)\Pi_{l=1}^{k}(a_i - l)\eta_k,$$

which follows because

$$\sum_{k=0}^{a_j+1} \Pi_{l=0}^{k-1}(a_j - l)\Pi_{l=1}^{k}(a_i - l)\eta_k - \sum_{k=0}^{a_j} \Pi_{l=0}^{k-1}(a_j - l)\Pi_{l=1}^{k}(a_i - l)\eta_k =$$

$$\Pi_{l=0}^{a_j}(a_j - l)\Pi_{l=1}^{a_j+1}(a_i - l)\eta_{a_j+1} = 0.$$

Inductively, it follows that

$$\sum_{k=0}^{a_j} \Pi_{l=0}^{k-1}(a_j - l)\Pi_{l=1}^{k}(a_i - l)\eta_k = \sum_{k=0}^{\infty} \Pi_{l=0}^{k-1}(a_j - l)\Pi_{l=1}^{k}(a_i - l)\eta_k.$$

Analogously,

$$\sum_{k=0}^{a_j-1} \Pi_{l=1}^{k}(a_j - l)\Pi_{l=0}^{k-1}(a_i - l)\eta_k = \sum_{k=0}^{\infty} \Pi_{l=1}^{k}(a_j - l)\Pi_{l=0}^{k-1}(a_i - l)\eta_k.$$

Therefore, we can extend the statement of Corollary 2 to read

$$\frac{\|G_{\mathbf{d_{-i}}}\|}{\|G_{\mathbf{d_{-j}}}\|} = \frac{a_i}{a_j}\left(\frac{\sum_{k=0}^{\infty} \Pi_{l=0}^{k-1}(a_j - l)\Pi_{l=1}^{k}(a_i - l)\eta_k}{\sum_{k=0}^{\infty} \Pi_{l=1}^{k}(a_j - l)\Pi_{l=0}^{k-1}(a_i - l)\eta_k}\right) \tag{3.9}$$

where η_k is defined in the statement of Corollary 2 for $k \leq \lfloor \frac{a_i+a_j-1}{2} \rfloor$ and $\eta_k = 1$ (or any finite number) for all other k. Expression (3.9) will be useful for providing flexibility in subsequent calculations.

The results we have proved so far apply to any graphic degree sequence. Under additional assumptions, at the cost of generality, we can go farther and obtain a power series representation for the ratio $\|G_{\mathbf{d_{-i}}}\|/\|G_{\mathbf{d_{-j}}}\|$. Specifically, if our graph is in some sense "sparse", equation (3.9) suggests that $\|G_{\mathbf{d_{-i}}}\|/\|G_{\mathbf{d_{-j}}}\| = a_i/a_j + O(\epsilon)$ for ϵ "small". We will elaborate on this idea in the following section.

3.2 COUNTING GRAPHS WITH DEGREE PRESERVING SWITCHES

We will use the idea of *degree preserving switches* to estimate the likelihood that two nodes have a common neighbor. This estimation will help us to derive a power series expansion of (3.9) that is valid if we let the number of nodes, or correspondingly the number of elements in each degree sequence, be sufficiently large. To obtain this result, it is important to set notation to specify how degrees behave as sequence length grows.

Definition 3. *Consider a sequence of degree sequences, $\{\mathbf{d}^N\}_{N\in\mathbb{N}}$, where each $\mathbf{d}^N \in \mathbb{Z}^{N\times 2}$. We say that $d_{max}(\{\mathbf{d}^N\}_{N\in\mathbb{N}}) = O(S^p)$, where $p \in \mathbb{R}$ and $S = S(\mathbf{d}^N)$ is the sum of the edges for the degree sequence $\mathbf{d}^N$, if and only if there exists a fixed constant $C \in \mathbb{R}$ such that $\overline{\lim}_{N\to\infty} \frac{max(\mathbf{d}^N)}{S^p} \leq C$, where the maximum is taken over all components of $\mathbf{d}^N$ and $\overline{\lim}$ denotes the limit supremum.*

For notational simplicity, we will omit explicit reference to the sequence of degree sequences when we write $d_{\max} = O(S^p)$. The use of the $O(S)$ notation is meant to indicate that we are not referring to the maximum degree of a fixed degree sequence. Analogously, we can say that for a sequence of degree sequences, the total number of nodes is $O(S^p)$ for some $p \in \mathbb{R}$. In addition, we say that the ith node has bounded (in)-degree in the limit of a sequence of degree sequences, if $\overline{\lim}_{N\to\infty}(\mathbf{a}^N)_i \leq C$, where $C \in \mathbb{R}$ and $(\mathbf{a}^N)_i$ denotes the ith element of the vector $\mathbf{a}^N$. At this juncture, we are ready to prove a result about the asymptotic likelihood of obtaining a graph, based on a uniform sampling of graphs realizing a degree sequence, in which two fixed nodes form edges with a common target node.

Theorem 8. *Consider a sequence of degree sequences such that $d_{max} = O(S^{\frac{1}{2}-\tau})$ for some $\tau > 0$. Pick an arbitrary node x and another node y such that y has a bounded number of edges in the limit of the sequence of degree sequences. In the limit of the sequence of degree sequences, the ratio of the number of realizations of graphs where there does not exist a node that receives an outward edge from (or supplies an inward to) both x and y relative to the number of graphs where exactly one node receives an outward edge from (or supplies an inward edge to) both x and y is $O(S^{2\tau})$.*

Proof. We will focus on the case of outward edges from x and y, as the inward case is anal-

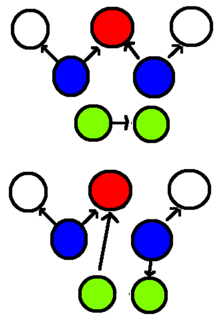

Figure 3.2.1: An example of the type of degree preserving switch described in the proof of Theorem 2. The top image is the input graph (i.e., the graph before the switch is applied) and the bottom is the output graph (generated by the switch). Note that in the output graph, all nodes have the same in and out degrees as in the input graph but the two blue nodes no longer have a common neighbor.

ogous. Consider a realization of a fixed degree sequence where a node (red in Figure 3.2.1) receives outward edges from nodes x and y (both blue). We can then define a degree preserving switch by choosing a remaining edge whose end points are not a red or blue node and redirecting it to the red node. Then to preserve the degree, we replace the edge that connects the right blue node (x) to the red node with an edge from x to the node that is now missing an edge. Unfortunately, this operation is not 1:1, since different input graphs can yield the same output graph, so we need to account for such repeats in order to accurately count the likelihood of a node receiving edges from both x and y.

Let us refer to the nodes receiving edges from a fixed node as the out-neighbors of that node and to the nodes supplying edges to a fixed node as the in-neighbors of that node. In the output graph of the degree preserving switch operator in Figure 3.2.1, denote the out-neighbors of node x by $\mathbf{S} = \{s_1, ..., s_l\}$ and the out-neighbors of node y by $\mathbf{T} = \{t_1, ..., t_m\}$, where l, m are the out-degrees of nodes x, y, respectively. Since the desired result only pertains to input graphs where x and y share exactly one common out-neighbor, we have $\mathbf{S} \cap \mathbf{T} = \emptyset$. Importantly, for each input graph that maps to this output graph, x must have

$l-1$ out-neighbors that are still its out-neighbors in the output graph, while the out-neighbors of y in the input and output graphs are the same.

Without loss of generality, suppose that in the input graph, t_1 is the node that has edges from both x and y. In the output graph denote the in-neighbors of t_1 as $\mathbf{U} = \{u_1, ..., u_n\}$ where n denotes the in-degree of t_1. In an input graph that is mapped by the degree preserving switch to the desired output graph, t_1 must have edges with $n-1$ of the nodes in U. Naturally, there are $\binom{n-1}{1}$ ways for this to happen. Since $l-1$ of the out-neighbors of x in the input graph must also be out-neighbors of x in the output graph, we conclude that there are fewer than $nl = O(S^{1-2\tau})$ ways to generate the same output if t_1 is the node that has edges with both x and y in the input graph.

We now repeat the same argument for each out-neighbor of y in the output graph. Since y has bounded degree, it follows that under the degree preserving operation, there are at most $O(S^{1-2\tau})$ ways to generate the same output.

Now that we know that our degree preserving operation is a $O(S^{1-2\tau}) : 1$ function, to finish off the proof, we need to identify how many degree preserving switches are possible from a single graph. The edges that are eligible to be switched connect nodes such that the source is not an in-neighbor of the common out-neighbor of x and y and the target does not already receive an edge from x (see Figure 3.2.1). Since x has degree at most $O(S^{\frac{1}{2}-\tau})$, that means the total number of edges corresponding to the neighbors of x is at most $O(S^{1-2\tau})$. Analogously, the common out-neighbor of x and y has degree at most $O(S^{\frac{1}{2}-\tau})$ and the number of edges corresponding to its in-neighbors is $O(S^{1-2\tau})$. Consequently, we have $O(S) - O(S^{1-2\tau}) = O(S)$ edges that we can choose from for the degree preserving switch operator to switch. Hence, for every graph where x and y have a common out-neighbor, the number of unique graphs where they do not have a common out-neighbor is at least the ratio of $O(S)$ to the number of input graphs that can map to each output graph, namely $O(S)/O(S^{1-2\tau}) = O(S^{2\tau})$, which is our desired result. □

Corollary 8. *Fix a sequence of degree sequences such that $d_{max} = O(S^{\frac{1}{2}-\tau})$. Let $k \in \mathbb{N}$. For any two nodes x and y, where y has bounded degree in the limit of the sequence of degree sequences, the ratio of the number of graphs where k nodes receive edges from both x and*

y compared to the number of graphs where $k+1$ nodes receive edges from both x and y is $O(S^{2\tau})$.

Proof. Perform the same switch technique as in Theorem 2 on a particular node that has an edge with both x and y, and the result follows analogously. □

Next, we apply Corollary 8 to begin to expand the terms in equation (3.4).

Corollary 9. *If $d_{max} = O(S^{\frac{1}{2}-\tau})$, then*

$$\|G_{\mathbf{d_{-i}}}\| = \frac{a_i}{a_j}\|G_{\mathbf{d_{-j}}}\|[1+O(S^{-2\tau})].$$

Proof. First suppose that either node i or node j has bounded degree. By Corollary 3, we know that for all $k \geq 1$,

$$\frac{\binom{a_i+a_j-2k-1}{a_j-k}\|G_{X_k}\|}{\binom{a_i+a_j-1}{a_j}\|G_{X_0}\|} = O(S^{-2k\tau}) \text{ and } \frac{\binom{a_i+a_j-2k-1}{a_j-k-1}\|G_{X_k}\|}{\binom{a_i+a_j-1}{a_j-1}\|G_{X_0}\|} = O(S^{-2k\tau}),$$

as these terms represent the number of graphs with k common neighbors of the two nodes i and j divided by the number of graphs where nodes i and j have no common neighbors. So we conclude from equation (3.6) in the proof of Corollary 2 that

$$\frac{\|G_{\mathbf{d_{-i}}}\|}{\|G_{\mathbf{d_{-j}}}\|} = \frac{\frac{\binom{a_i+a_j-1}{a_j}}{\binom{a_i+a_j-1}{a_j-1}}+\sum_{k=1}^{a_j}\frac{\binom{a_i+a_j-2k-1}{a_j-k}}{\binom{a_i+a_j-1}{a_j-1}}\frac{\|G_{X_k}\|}{\|G_{X_0}\|}}{1+\sum_{k=1}^{a_j-1}\frac{\binom{a_i+a_j-2k-1}{a_j-k-1}}{\binom{a_i+a_j-1}{a_j-1}}\frac{\|G_{X_k}\|}{\|G_{X_0}\|}} = \frac{\frac{\binom{a_i+a_j-1}{a_j}}{\binom{a_i+a_j-1}{a_j-1}}+\sum_{k=1}^{a_j}\frac{\binom{a_i+a_j-2k-1}{a_j-k}}{\binom{a_i+a_j-1}{a_j-1}}\frac{\|G_{X_k}\|}{\|G_{X_0}\|}}{1+O(S^{-2\tau})}$$

$$= \frac{\binom{a_i+a_j-1}{a_j}}{\binom{a_i+a_j-1}{a_j-1}}\left[\frac{1+\sum_{k=1}^{a_j}\frac{\binom{a_i+a_j-2k-1}{a_j-k}}{\binom{a_i+a_j-1}{a_j}}\frac{\|G_{X_k}\|}{\|G_{X_0}\|}}{1+O(S^{-2\tau})}\right] = \frac{a_i}{a_j}\left[\frac{1+O(S^{-2\tau})}{1+O(S^{-2\tau})}\right] = \frac{a_i}{a_j}[1+O(S^{-2\tau})].$$

To extend this relationship to the general case where nodes i and j both have degree $O(S^{\frac{1}{2}-\tau})$, we consider the degree sequences $\mathbf{d^{(i)}} = (\{\mathbf{a},1\}-\mathbf{e_i},\{\mathbf{b},0\}), \mathbf{d^{(j)}} = (\{\mathbf{a},1\}-\mathbf{e_j},\{\mathbf{b},0\}) \in \mathbb{Z}^{(N+1)\times 2}$, for which

$$\frac{\|G_{\mathbf{d_{-i}}}\|}{\|G_{\mathbf{d_{-j}}}\|} = \frac{\|G_{\mathbf{d^{(i)}_{-(N+1)}}}\|}{\|G_{\mathbf{d^{(i)}_{-j}}}\|}\frac{\|G_{\mathbf{d^{(j)}_{-i}}}\|}{\|G_{\mathbf{d^{(j)}_{-(N+1)}}}\|}, \tag{3.10}$$

where node $N+1$ has bounded degree and $\mathbf{d}_{-\mathbf{i}}^{(\mathbf{j})} = \mathbf{d}_{-\mathbf{j}}^{(\mathbf{i})}$. Consequently,

$$\frac{\|G_{\mathbf{d}_{-\mathbf{i}}}\|}{\|G_{\mathbf{d}_{-\mathbf{j}}}\|} = \frac{a_i}{a_j}[1 + O(S^{-2\tau})].$$

□

We conclude this section noting that the proof technique can be extended to more general cases. We can relax the constraint that $d_{max} = O(S^{\frac{1}{2}-\tau})$ and still attain that two nodes in general should not share common neighbors. For more details we refer the reader to Section 3.7. On a similar note, the notion that a node of bounded degree and an arbitrary node should not share common neighbors extends to cases beyond directed graphs with loops, including directed and undirected graphs without loops. We discuss in Section 3.7 how to extend these arguments to these cases, noting that similar ideas also apply to graphs with other edges prohibited besides loops. Section 3.7 explains how to iteratively attain more refined approximations from the relatively crude approximation given by Corollary 9. These results also generalize in ways that are discussed in the appendices.

3.3 ASYMPTOTIC ENUMERATION TO ARBITRARY ORDERS OF ACCURACY

Since we have established some fundamental results in the prior section, we can now derive our general asymptotic enumeration.

We start by establishing an alternative expression for $\|G_{\mathbf{d}_{-\mathbf{i}}}\|/\|G_{\mathbf{d}_{-\mathbf{j}}}\|$. For this statement, define X_{0_i} as the set of all residual degree sequences $(\mathbf{a} - a_i\mathbf{e_i} - \mathbf{e_j}, \mathbf{b} - \mathbf{s})$ with $\mathbf{s}$ constructed by removing one (outgoing) edge from each of a_i nodes; define X_{1_i} as the set of all residual degree sequences constructed by removing two outgoing edges from one node and one edge from $a_i - 2$ nodes; and define X_{0_j},X_{1_j} analogously from $(\mathbf{a} - \mathbf{e_i} - a_j\mathbf{e_j}, \mathbf{b} - \mathbf{s})$.

Corollary 10.

$$\frac{\|G_{d_{-i}}\|}{\|G_{d_{-j}}\|} = \frac{a_i}{a_j}\left(\frac{1 + \|G_{X_{1_i}}\|/(a_i\|G_{X_{0_i}}\|)}{1 + \|G_{X_{1_j}}\|/(a_j\|G_{X_{0_j}}\|)}\right) \tag{3.11}$$

Proof. The proof is in the spirit of Corollary 4. Let $\mathbf{d^{(i)}} = (\{\mathbf{a}, 1\} - \mathbf{e_i}, \{\mathbf{b}, 0\}), \mathbf{d^{(j)}} = (\{\mathbf{a}, 1\} - \mathbf{e_j}, \{\mathbf{b}, 0\}) \in \mathbb{Z}^{(N+1)\times 2}$. We previously established equation (3.10). But applying Theorem 1 yields

$$\frac{\|G_{\mathbf{d^{(i)}_{-(N+1)}}}\|}{\|G_{\mathbf{d^{(i)}_{-j}}}\|} = \frac{\|G_{X_{0_j}}\|}{a_j\|G_{X_{0_j}}\| + \|G_{X_{1_j}}\|} = \frac{1}{a_j}\left(\frac{1}{1 + \|G_{X_{1_j}}\|/(a_j\|G_{X_{0_j}}\|)}\right).$$

Similarly,

$$\frac{\|G_{\mathbf{d^{(j)}_{-i}}}\|}{\|G_{\mathbf{d^{(j)}_{-(N+1)}}}\|} = \frac{a_i\|G_{X_{0_i}}\| + \|G_{X_{1_i}}\|}{\|G_{X_{0_i}}\|} = a_i(1 + \frac{\|G_{X_{1_i}}\|}{a_i\|G_{X_{0_i}}\|}).$$

□

Corollary 10 will be useful to us when combined with the observation that for an arbitrary degree sequence $\mathbf{m}$,

$$\frac{\|G_{X_{1_i}}\|}{\|G_{X_{0_i}}\|} = \frac{\sum_{\mathbf{x}\in X_{1_i}} \|G_{\mathbf{x}}\|}{\sum_{\mathbf{x}\in X_{0_i}} \|G_{\mathbf{x}}\|} = \frac{\sum_{\mathbf{x}\in X_{1_i}} \|G_{\mathbf{x}}\|/\|G_{\mathbf{m}}\|}{\sum_{\mathbf{x}\in X_{0_i}} \|G_{\mathbf{x}}\|/\|G_{\mathbf{m}}\|}. \tag{3.12}$$

That is, equation (3.11) represents a recursion that expresses the ratio of the number of graphs of two different degree sequences as a function of the ratios of the numbers of graphs of various other degree sequences.

We now state the first of three theorems that will enable us to reach the desired asymptotic enumeration results of arbitrary order.

Theorem 9. *Let* $\mathbf{d} = (\mathbf{a}, \mathbf{b})$ *with* $\sum a_n = \sum b_n \pm 1$ *and* $d_{max} = O(S^{\frac{1}{2}-\tau})$*. Consider an approximation that satisfies the equation*

$$\frac{\|G_{\mathbf{d_{-i}}}\|}{\|G_{\mathbf{d_{-j}}}\|} = f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}, \sigma)(1 + O(S^{-2w\tau}))$$

where $w \geq 1$, $2w\tau \leq \gamma$ *(we define* γ *below), and the last argument* σ *either equals* $\mathbf{a}$*, which denotes that* $\sum a_n = \sum b_n + 1$ *and the two degree sequences in the ratio differ in their in-degree sequences (i.e., the in-degree sequence is being used to define* $\mathbf{d_{-i}}$*), or* σ *equals* $\mathbf{b}$*, which has the analogous connotation with respect to out-degree sequences, with* $\sum a_n = \sum b_n - 1$*. Assume that* $f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}, \sigma) = h(\mathbf{e}_i, \mathbf{d}, \sigma)/h(\mathbf{e}_j, \mathbf{d}, \sigma)$ *for some function* h*. Furthermore, suppose that for* $m = O(S^{\frac{1}{2}-\tau})$,

$$h(\mathbf{e}_i, \mathbf{d}_1, \sigma_1) = h(\mathbf{e}_i, \mathbf{d}_0, \sigma_0)(1 + O(S^{-\gamma})) \tag{3.13}$$

where $\|\mathbf{d}_1 - \mathbf{d}_0\|_1 \leq m$*,* $\|\mathbf{d}_1 - \mathbf{d}_0\|_\infty \leq 1$*,* σ_i *is either* $\mathbf{a}_i$ *or* $\mathbf{b}_i$ *(i.e., the in- or out-degree sequence of* $\mathbf{d}_i$*), and the following equalities of dot products hold:* $\sigma_1 \cdot \mathbf{e}_i = \sigma_0 \cdot \mathbf{e}_i$*,* $\sigma_1 \cdot \mathbf{e}_j = \sigma_0 \cdot \mathbf{e}_j$*. If* $\sigma = \mathbf{a}$*, then there exists a sharper approximation*

$$\frac{\|G_{\mathbf{d_{-i}}}\|}{\|G_{\mathbf{d_{-j}}}\|} = g(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}, \mathbf{a})(1 + O(S^{-2(w+1)\tau}))$$

where

$$g(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}, \mathbf{a}) = \frac{a_i}{a_j} \exp(\log(1 + \frac{(a_i - 1)\sum_{x_1 \neq \ldots \neq x_{a_i - 1} = x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, \mathbf{d}_{k,i,j}, \mathbf{b}_k)}{\sum_{x_1 \neq \ldots \neq x_{a_i - 1} \neq x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, \mathbf{d}_{k,i,j}, \mathbf{b})}) -$$

$$\log(1 + \frac{(a_j - 1)\sum_{x_1 \neq \ldots \neq x_{a_j - 1} = x_{a_j}} \Pi_{k=1}^{a_j} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, \mathbf{d}_{k,j,i}, \mathbf{b})}{\sum_{x_1 \neq \ldots \neq x_{a_j - 1} \neq x_{a_j}} \Pi_{k=1}^{a_j} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, \mathbf{d}_{k,j,i}, \mathbf{b})}))$$

for any arbitrary choice of indices $\{u_k\}$*, where* $\mathbf{d}_{k,i,j} = (\mathbf{a} - a_i\mathbf{e}_i - \mathbf{e}_j, \mathbf{b} - \sum_{j=1}^{a_i - k + 1} \mathbf{e}_{u_j} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j})$*. A similar sharpened approximation, with* g *depending on* $\mathbf{b}$*, holds if* $\sigma = \mathbf{b}$*.*

We will postpone the motivation for the assumptions that $f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}, \sigma) = h(\mathbf{e}_i, \mathbf{d}, \sigma)/h(\mathbf{e}_j, \mathbf{d}, \sigma)$ for some function h and that perturbing the third argument of f (the degree sequence) only results in relatively small changes in h until Section 6. In Section 5, when we use the above theorem to explicitly compute asymptotics of the ratio $\|G_{\mathbf{d_{-i}}}\|/\|G_{\mathbf{d_{-j}}}\|$, we will be able to verify these assumptions directly. Even at this stage, we already know that the approximation $\|G_{\mathbf{d_{-i}}}\|/\|G_{\mathbf{d_{-j}}}\| = \frac{a_i}{a_j}(1 + O(S^{-2\tau}))$ does not depend on the degrees of the nodes in the degree sequence other than nodes i and j. Similarly, this approximation can be expressed in terms of a decomposition like that assumed in Theorem 9, given by $\|G_{\mathbf{d_{-i}}}\|/\|G_{\mathbf{d_{-j}}}\| = [h(\mathbf{e}_i, \mathbf{d}, \mathbf{a})/h(\mathbf{e}_j, \mathbf{d}, \mathbf{a})](1 + O(S^{-2\tau}))$, where $h(\mathbf{e}_i, \mathbf{d}, \mathbf{a}) = a_i$. We now proceed with the proof.

Proof. Consider $\mathbf{d}$ such that $\sum a_n = \sum b_n + 1$. We know from equation (3.11) that

$$\frac{\|G_{\mathbf{d_{-i}}}\|}{\|G_{\mathbf{d_{-j}}}\|} = \frac{a_i}{a_j} \exp(\log(1 + \frac{\|G_{X_{1_i}}\|}{a_i\|G_{X_{0_i}}\|}) - \log(1 + \frac{\|G_{X_{1_j}}\|}{a_j\|G_{X_{0_j}}\|})).$$

Let

$$\|G_{\mathbf{d_{-i}}}\|/\|G_{\mathbf{d_{-j}}}\| = f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}, \mathbf{a})(1 + O(S^{-2w\tau}))$$

for some $w \geq 1$. Our goal is to show that substituting the decomposition for f into equation (3.30) yields a sharper approximation g, as specified in the theorem statement (the proof with $\sigma = \mathbf{b}$ is analogous).

To prove this claim rigorously, we show that using f to approximate $\frac{\|G_{X_{1_i}}\|}{a_i\|G_{X_{0_i}}\|}$ can allow us to derive an improved approximation of $\frac{\|G_{d_{-i}}\|}{\|G_{d_{-j}}\|}$, in a sense that will be made precise later. Since $\frac{\|G_{X_{1_i}}\|}{a_i\|G_{X_{0_i}}\|}$ is equivalent to $\frac{\|G_{X_{1_j}}\|}{a_j\|G_{X_{0_j}}\|}$, we can carry over our results to the latter expression to complete the derivation.

For any $\mathbf{u} \in X_{0_i}$, equation (3.12) yields

$$\frac{\|G_{X_{1_i}}\|}{a_i\|G_{X_{0_i}}\|} = \frac{\sum_{\mathbf{x}\in X_{1_i}} \|G_{\mathbf{x}}\|/\|G_{\mathbf{u}}\|}{a_i \sum_{\mathbf{x}\in X_{0_i}} \|G_{\mathbf{x}}\|/\|G_{\mathbf{u}}\|}. \tag{3.14}$$

Now, for any $\mathbf{x} \in X_{0_i} \cup X_{1_i}$, the bidegree sequences $\mathbf{u}$ and $\mathbf{x}$ include the same number of edges and identical in-degree sequences. By definition we can write $\mathbf{x} = (\mathbf{a} - a_i\mathbf{e}_i - \mathbf{e}_j, \mathbf{b} - \sum_{j=1}^{a_i} \mathbf{e}_{x_j})$ and $\mathbf{u} = (\mathbf{a} - a_i\mathbf{e}_i - \mathbf{e}_j, \mathbf{b} - \sum_{j=1}^{a_i} \mathbf{e}_{u_j})$. Let $\mathbf{d}_1 = \mathbf{u}, \mathbf{d}_{a_{i+1}} = \mathbf{x}$, and define intermediate degree sequences $\mathbf{d}_k = (\mathbf{a} - a_i\mathbf{e}_i - \mathbf{e}_j, \mathbf{b} - \sum_{j=1}^{a_i-k+1} \mathbf{e}_{u_j} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j})$ for each $k = 2, \ldots, a_i$. Note that the sum of the in-degrees equals the sum of the out-degrees in each $\mathbf{d}_k$, since by assumption, $\sum a_n = \sum b_n + 1$.

Letting

$$\phi(\mathbf{e_i}, \mathbf{e_j}, \mathbf{d}, \sigma) := \|G_{\mathbf{d_{-i}}}\|/\|G_{\mathbf{d_{-j}}}\| \tag{3.15}$$

for any choice of $\mathbf{d}$, these definitions imply that for our particular $\mathbf{d}$ under consideration,

$$\frac{\|G_{\mathbf{x}}\|}{\|G_{\mathbf{u}}\|} = \Pi_{k=1}^{a_i} \frac{\|G_{\mathbf{d}_{k+1}}\|}{\|G_{\mathbf{d}_k}\|} = \Pi_{k=1}^{a_i} \phi(\mathbf{e}_{x_k}, \mathbf{e}_{u_{a_i-k+1}}, (\mathbf{a} - a_i\mathbf{e}_i - \mathbf{e}_j, \mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} - \sum_{j=1}^{a_i-k} \mathbf{e}_{u_j}, \mathbf{b}), \tag{3.16}$$

where now the sum of the in-degrees in the third argument of ϕ in (3.15) is *one less than* the sum of the out-degrees. Recall that the final argument $\mathbf{b}$ of ϕ in (3.15) implies that in the numerator and denominator of the right hand side of (3.15), a degree is being subtracted off of the *out-degree sequence* of the bidegree sequence given in the third argument of ϕ, which means that the in- and out-degree sequences in $\mathbf{d_{-i}}, \mathbf{d_{-j}}$ end up with the same sums. The particular components of the out-degree sequence from which a degree is being subtracted are specified in the first and second arguments of ϕ for the numerator and denominator, respectively.

Now, define Δ_i to be the difference between $\frac{\|G_{X_{1_i}}\|}{a_i\|G_{X_{0_i}}\|}$ evaluated using the exact ratio ϕ and the same quantity evaluated using the approximation f. Recall that each $\mathbf{x} \in X_{0_i}$ was defined by removing the a_i incoming edges to node i along with one outgoing edge from each of a_i distinct nodes. The number of resulting bidegree sequences is the same as the number of bidegree sequences in which the a_i outgoing edges are directed to node i instead of being removed. There is an analogous equivalence for each $\mathbf{x} \in X_{1_i}$. Hence, we can write Δ_i as

$$\Delta_i =$$

$$\frac{(a_i-1)\sum_{x_1\neq\ldots\neq x_{a_{i-1}}=x_{a_i}} \Pi_{k=1}^{a_i}\phi(\mathbf{e}_{x_k},\mathbf{e}_{u_{a_i-k+1}},(\mathbf{a}-a_i\mathbf{e}_i-\mathbf{e}_j,\mathbf{b}-\sum_{j=1}^{k-1}\mathbf{e}_{x_j}-\sum_{j=1}^{a_i-k}\mathbf{e}_{u_j}),\mathbf{b})}{\sum_{x_1\neq\ldots\neq x_{a_{i-1}}\neq x_{a_i}} \Pi_{k=1}^{a_i}\phi(\mathbf{e}_{x_k},\mathbf{e}_{u_{a_i-k+1}},(\mathbf{a}-a_i\mathbf{e}_i-\mathbf{e}_j,\mathbf{b}-\sum_{j=1}^{k-1}\mathbf{e}_{x_j}-\sum_{j=1}^{a_i-k}\mathbf{e}_{u_j}),\mathbf{b})}-$$

$$\frac{(a_i-1)\sum_{x_1\neq\ldots\neq x_{a_{i-1}}=x_{a_i}} \Pi_{k=1}^{a_i}f(\mathbf{e}_{x_k},\mathbf{e}_{u_{a_i-k+1}},(\mathbf{a}-a_i\mathbf{e}_i-\mathbf{e}_j,(\mathbf{b}-\sum_{j=1}^{k-1}\mathbf{e}_{x_j}-\sum_{j=1}^{a_i-k}\mathbf{e}_{u_j}),\mathbf{b})}{\sum_{x_1\neq\ldots\neq x_{a_{i-1}}\neq x_{a_i}} \Pi_{k=1}^{a_i}f(\mathbf{e}_{x_k},\mathbf{e}_{u_{a_i-k+1}},(\mathbf{a}-a_i\mathbf{e}_i-\mathbf{e}_j,(\mathbf{b}-\sum_{j=1}^{k-1}\mathbf{e}_{x_j}-\sum_{j=1}^{a_i-k}\mathbf{e}_{u_j}),\mathbf{b})}. \tag{3.17}$$

Since the choice for u_{a_i-k+1} is arbitrary and assumption (3.13) in the theorem statement holds, we can simplify the notation by using $f_k(x_k)$ in place of the full expression for f. (We will defer a technical point regarding this simplification to the end of the proof.) In contrast , ϕ does depend on the degree sequence and on .$x_1, ..., x_{k-1}$. But for simplicity, we abuse notation and write $\phi_k(x_k)$. This reduces to a more tractable (but slightly misleading) notation:

$$\Delta_i = \frac{(a_i-1)\sum_{x_1\neq\ldots\neq x_{a_{i-1}}=x_{a_i}} \Pi_{k=1}^{a_i}\phi_k(x_k)}{\sum_{x_1\neq\ldots\neq x_{a_{i-1}}\neq x_{a_i}} \Pi_{k=1}^{a_i}\phi_k(x_k)} - \frac{(a_i-1)\sum_{x_1\neq\ldots\neq x_{a_{i-1}}=x_{a_i}} \Pi_{k=1}^{a_i}f_k(x_k)}{\sum_{x_1\neq\ldots\neq x_{a_{i-1}}\neq x_{a_i}} \Pi_{k=1}^{a_i}f_k(x_k)}.$$

Denote D_0 as the set of sets of a_i distinct indices in $\{1,\ldots,N\}$ and D_1 as the set of sets of a_i indices in $\{1,\ldots,N\}$ such that the first a_i-2 are distinct and the final two are equal. Writing Δ_i as a single fraction, we obtain

$$\Delta_i = \frac{(a_i-1)[\sum_{D_1}\Pi_{k=1}^{a_i}\phi_k(x_k)\sum_{D_0}\Pi_{k=1}^{a_i}f_k(x_k) - \sum_{D_1}\Pi_{k=1}^{a_i}f_k(x_k)\sum_{D_0}\Pi_{k=1}^{a_i}\phi_k(x_k)]}{\sum_{D_0}\Pi_{k=1}^{a_i}\phi_k(x_k)\sum_{D_0}\Pi_{k=1}^{a_i}f_k(x_k)}$$
$$= \frac{(a_i-1)[\sum_{D_1,D_0}(\Pi_{x_k\in D_1}\phi_k(x_k)\Pi_{x_k\in D_0}f_k(x_k) - \Pi_{x_k\in D_1}f_k(x_k)\Pi_{x_k\in D_0}\phi_k(x_k))]}{\sum_{D_0}\Pi_{k=1}^{a_i}\phi_k(x_k)\sum_{D_0}\Pi_{k=1}^{a_i}f_k(x_k)}.$$

We now write $\phi_k = f_k(1+\xi_k)$ where ξ_k depends only on $x_1, ..., x_k$ (but we omit the dependence) and $\xi_k = O(S^{-2w\tau})$ from the definition of f. Furthermore, denote $\delta_k = 0$ if $k = a_i$ or $k = a_i - 1$ and $\delta_k = 1$ otherwise. These steps yield

$\Delta_i =$

$$\frac{(a_i-1)[\sum_{D_1,D_0}\Pi_{x_k\in D_1}f_k(x_k)(1+\xi_k\delta_k)\Pi_{x_k\in D_0}f_k(x_k) -}{\sum_{D_0}\Pi_{k=1}^{a_i}\phi_k(x_k)\sum_{D_0}\Pi_{k=1}^{a_i}f_k(x_k)} \tag{3.18}$$

$$\frac{(a_i-1)[\sum_{D_1,D_0}(\Pi_{x_k\in D_1}f_k(x_k)\Pi_{x_k\in D_0}f_k(x_k)(1+\xi_k\delta_k)) + \epsilon]}{\sum_{D_0}\Pi_{k=1}^{a_i}\phi_k(x_k)\sum_{D_0}\Pi_{k=1}^{a_i}f_k(x_k)} \tag{3.19}$$

where ϵ is the compensatory term for zeroing out certain terms by inserting the δ_k into equation (3.18), which we can express as $\epsilon = \epsilon_1 + \epsilon_2 - \epsilon_3 - \epsilon_4$ for

$$\epsilon_1 = \sum_{D_1}\xi_{a_i}f_{a_i}(x_{a_i})f_{a_i-1}(x_{a_i-1})\Pi_{k\neq a_i-1,a_i}f_k(x_k)(1+\xi_k)\sum_{D_0}\Pi_{x_k\in D_0}f_k(x_k), \tag{3.20}$$

$$\epsilon_2 = \sum_{D_1}\xi_{a_i-1}f_{a_i-1}(x_{a_i-1})\Pi_{k\neq a_i-1}f_k(x_k)(1+\xi_k)\sum_{D_0}\Pi_{x_k\in D_0}f_k(x_k), \tag{3.21}$$

$$\epsilon_3 = \sum_{D_1}\Pi_{x_k\in D_1}f_k(x_k)\sum_{D_0}\xi_{a_i-1}f_{a_i}(x_{a_i})f_{a_i-1}(x_{a_i-1})\Pi_{k\neq a_i-1,a_i}f_k(x_k)(1+\xi_k), \tag{3.22}$$

$$\epsilon_4 = \sum_{D_1}\Pi_{x_k\in D_1}f_k(x_k)\sum_{D_0}\xi_{a_i}f_{a_i}(x_{a_i})\Pi_{k\neq a_i}f_k(x_k)(1+\xi_k). \tag{3.23}$$

Now, for $k = a_i$ or $k = a_i - 1$, $f_k(x_k)(1 + \xi_k\delta_k) = f_k(x_k)$ by definition. Factoring $f_k(x_k)$ out for those choices of x_k in both D_0 and D_1, applying a version of the mean value theorem and using the fact that ξ_k only depends on $x_1, ..., x_k$ enables us to integrate out the last two variables of D_0 and D_1. Since the first $a_i - 2$ indices are distinct in each element of both D_0 and D_1, if we define D_* as the set of sets of $a_i - 2$ distinct indices, then the expression for Δ_i can now be written as

$$\Delta_i = \tag{3.24}$$

$$\frac{(a_i - 1)[\lambda\{\sum_{D_*,D_*} (\Pi_{x_k \in D_*} f_k(x_k)(1+\xi_k)\Pi_{x_k \in D_*} f_k(x_k) - \Pi_{x_k \in D_*} f_k(x_k)\Pi_{x_k \in D_*} f_k(x_k)(1+\xi_k))\}]}{\sum_{D_0} \Pi_{k=1}^{a_i} \phi_k(x_k) \sum_{D_0} \Pi_{k=1}^{a_i} f_k(x_k)} + \tag{3.25}$$

$$\frac{(a_i - 1)\epsilon}{\sum_{D_0} \Pi_{k=1}^{a_i} \phi_k(x_k) \sum_{D_0} \Pi_{k=1}^{a_i} f_k(x_k)} = \tag{3.26}$$

$$\frac{(a_i - 1)[\epsilon]}{\sum_{D_0} \Pi_{k=1}^{a_i} \phi_k(x_k) \sum_{D_0} \Pi_{k=1}^{a_i} f_k(x_k)} \tag{3.27}$$

where λ is the constant from the application of the mean value theorem.

To bound Δ_i, we use the crude approximation that for $k = a_i - 1, a_i$, both $f_k(x_k, \mathbf{b}) = \frac{\mathbf{b}_{x_k}}{\mathbf{b}_{u_k}}(1 + O(S^{-2\tau}))$ and $\phi_k(x_k) = \frac{b_{x_k}}{b_{u_k}}(1 + O(S^{-2\tau}))$, and we multiply the numerator and denominator of Δ_i by $(\Pi_{k=a_i-1}^{a_i} b_{u_k})^2$. From equation (4.16), we have that

$$(a_i - 1)(\Pi_{k=a_i-1}^{a_i} b_{u_k})^2 \epsilon_1 \leq$$

$$(a_i - 1) \sum_{D_1} (\Pi_{k=a_i-1}^{a_i} b_{u_k}) \xi_{a_i} f_{a_i}(x_{a_i}) f_{a_i-1}(x_{a_i-1}) \Pi_{k \neq a_i-1, a_i} f_k(x_k)(1+\xi_k) \sum_{D_0} (\Pi_{k=a_i-1}^{a_i} b_{u_k}) \Pi_{x_k \in D_0} f_k(x_k).$$

Note that $\xi_{a_i} = O(S^{-2w\tau})$ and $a_i \leq d_{max}$. Moreover, using the relationship that $\sum_{m=1}^{N} b_{u_k}\phi_k(x_m, \mathbf{b}) = S(1 + O(S^{-2\tau}))$ to integrate out the last two variables x_{a_i-1} and x_{a_i} from D_0, we can obtain the bound

$$\sum_{D_0}(\Pi_{k=a_i-1}^{a_i} b_{u_k})\Pi_{x_k\in D_0}^{a_i} f_k(x_k) \leq \sum_{D_0*}(S + O(S^{1-2\tau}))^2 \Pi_{x_k\in D_0*} f_k(x_k).$$

Hence,

$$(a_i - 1)(\Pi_{k=a_i-1}^{a_i} b_{u_k})^2 \epsilon_1 \leq$$

$$d_{max} O(S^{-2w\tau}) \sum_{D_1}(\Pi_{k=a_i-1}^{a_i} b_{u_k}) f_{a_i}(x_{a_i}) f_{a_i-1}(x_{a_i-1}) \Pi_{k\neq a_i-1,a_i} f_k(x_k)(1+\xi_k) \sum_{D_0*}(S + O(S^{1-2\tau}))^2 \Pi_{x_k\in D_0*} f_k(x_k).$$

Similarly, we conclude that

$$(a_i - 1)(\Pi_{k=a_i-1}^{a_i} b_{u_k})^2 \epsilon_1 \leq O(S^{-2w\tau}) d_{max}^2 [S + O(S^{1-2\tau})]^3 \sum_{D_1*} \Pi_{x_k\in D_1*} f_k(x_k) \sum_{D_0*} \Pi_{x_k\in D_0*} f_k(x_k)$$

by replacing the summation over D_1 with a summation over D_{1*}, again integrating out the last two variables x_{a_i-1} and x_{a_i} from D_1 using $b_{u_k}\phi_k(x_m, \mathbf{b}) = b_{x_m}(1 + O(S^{-2\tau}))$. Repeating this argument for all ϵ_i $i = 2, 3, 4$, we obtain that

$$(a_i - 1)(\Pi_{k=a_i-1}^{a_i} b_{u_k})^2 \epsilon \leq 4 d_{max}^2 (S + O(S^{1-2\tau}))^3 O(S^{-2w\tau}) \sum_{D_1*} \Pi_{x_k\in D_1*} f_k(x_k)(1+\xi_k) \sum_{D_0*} \Pi_{x_k\in D_0*} f_k(x_k).$$

Similarly,

$$(\Pi_{k=a_i-1,a_i} b_{u_k})^2 \sum_{D_0} \Pi_{k=1}^{a_i} \phi_k(x_k) \sum_{D_0} \Pi_{k=1}^{a_i} f_k(x_k) \geq$$

$$(S - O(S^{1-2\tau}))^4 \sum_{D_1*} \Pi_{x_k\in D_1*} f_k(x_k)(1+\xi_k) \sum_{D_0*} \Pi_{x_k\in D_0*} f_k(x_k).$$

Simplifying, using the fact that $d_{max} = O(S^{\frac{1}{2}-\tau})$, we obtain

$$\Delta_i = \frac{(a_i - 1)(\Pi_{k=a_i-1}^{a_i} b_{u_k})^2 \epsilon}{(\Pi_{k=a_i-1,a_i} b_{u_k})^2 \sum_{D_0} \Pi_{k=1}^{a_i} \phi_k(x_k) \sum_{D_0} \Pi_{k=1}^{a_i} f_k(x_k)} = O(S^{-2(w+1)\tau}).$$

Therefore, we conclude that g as defined in the theorem statement can be approximated as $g(\mathbf{e}_i, \mathbf{e_j}, \mathbf{d}, \mathbf{a}) = \frac{\|G_{\mathbf{d_{-i}}}\|}{\|G_{\mathbf{d_{-j}}}\|}(1 + O(S^{-2(w+1)\tau}))$, even though our approximation for f is $f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}, \mathbf{x}) = \frac{\|G_{\mathbf{d_{-i}}}\|}{\|G_{\mathbf{d_{-j}}}\|}(1 + O(S^{-2w\tau}))$.

We have nearly completed the proof. We argued above that the choice of u_k should not have any impact in evaluating expressions such as

$$(a_i - 1) \sum_{x_1 \neq \ldots \neq x_{a_{i-1}} = x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, (\mathbf{a} - a_i \mathbf{e}_i, \mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} - \sum_{j=1}^{a_i - k} \mathbf{e}_{u_j}), \mathbf{b}) = \sum_{\mathbf{x} \in X_{1_i}} \frac{\|G_{\mathbf{x}}\|}{\|G_{\mathbf{u}}\|}$$

By our assumptions on f, the dependence on the degree sequence is ignored (i.e., represents a higher order term) unless one of the $u_{a_i - m+1}$ equals one of the x_k's. (That is, when we evaluate $\Pi_k f(\mathbf{e}_{x_k}, \mathbf{e}_{u_{a_i-k+1}}, (\mathbf{a} - a_i \mathbf{e}_i, \mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} - \sum_{j=1}^{a_i-k} \mathbf{e}_{u_j}), \mathbf{b})$ the out-degree of node x_k may be $b_{x_k} - 1$ and not b_{x_k}.)

If this is the case, then consider the product

$$f(\mathbf{e}_{x_k}, \mathbf{e}_{u_{a_i-k+1}}, \mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} - \sum_{j=1}^{a_i-k} \mathbf{e}_{u_j}) f(\mathbf{e}_{x_m}, \mathbf{e}_{u_{a_i-m+1}}, \mathbf{b} - \sum_{j=1}^{m-1} \mathbf{e}_{x_j} - \sum_{j=1}^{a_i-m} \mathbf{e}_{u_j}) \tag{3.28}$$

where we omitted the dependence on the in-degree sequence and the last argument $\mathbf{b}$. Now when evaluating $f(\mathbf{e}_{x_k}, \mathbf{e}_{u_{a_i-k+1}}, \mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} - \sum_{j=1}^{a_i-k} \mathbf{e}_{u_j})$, the degree of the x_k is $b_{x_k} - 1$ if and only if $a_i - m + 1 \leq a_i - k$, so in both terms of the product the out-degree of node x_k is $b_{x_k} - 1$.

We will now invoke assumption (3.13). We will condense notation a bit: in the function h, if $\mathbf{d} = (\mathbf{a}, \mathbf{b})$ and σ selects the in-degree (out-degree) sequence, then we write $h(\mathbf{e}_i, \mathbf{a})$ ($h(\mathbf{e}_i, \mathbf{b})$). We can thus write (3.28) as

$$\left(\frac{h(x_k, \mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} - \sum_{j=1}^{a_i-k} \mathbf{e}_{u_j})}{h(u_{a_i-k+1}, \mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} - \sum_{j=1}^{a_i-k} \mathbf{e}_{u_j})} \right) \left(\frac{h(x_m, \mathbf{b} - \sum_{j=1}^{m-1} \mathbf{e}_{x_j} - \sum_{j=1}^{a_i-m} \mathbf{e}_{u_j})}{h(u_{a_i-m+1}, \mathbf{b} - \sum_{j=1}^{m-1} \mathbf{e}_{x_j} - \sum_{j=1}^{a_i-m} \mathbf{e}_{u_j})} \right).$$

But since $x_k = u_{a_i-m+1}$,

$$\frac{h(x_k, \mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} - \sum_{j=1}^{a_i-k} \mathbf{e}_{u_j})}{h(u_{a_i-m+1}, \mathbf{b} - \sum_{j=1}^{m-1} \mathbf{e}_{x_j} - \sum_{j=1}^{a_i-m} \mathbf{e}_{u_j})} = (1 + O(S^{-2w\tau})) = \frac{h(x_k, \mathbf{z_1})}{h(u_{a_i-m+1}, \mathbf{z_2})}(1 + O(S^{-2w\tau}))$$

for any arbitrary vectors $\mathbf{z_1}$,$\mathbf{z_2}$ that satisfy assumption (3.13), as $x_k = u_{a_i-m+1}$. In particular, we can choose $\mathbf{z_1}$,$\mathbf{z_2}$ such that the out-degree of node x_k is b_{x_k} and not $b_{x_k} - 1$,hence the dependence on the degree sequence can be ignored as initially claimed. □

Even with the many assumptions in the statement of Theorem 9, we still do not have our asymptotic enumeration result for counting the number of graphs realizing a given bidegree sequence. The issue is that we would need to evaluate products of approximations $\Pi_{i=1}^{a_i} f(x_i)(1 + O(S^{-2w\tau}))$ where $a_i = O(S^{\frac{1}{2}-\tau})$, which diverges as $S \to \infty$. To avoid this problem, we note that Theorem 9 does give us a way to shrink the error term in the product, decreasing the power of S by 1 in each step. Thus, by repeatedly applying Theorem 9, we can obtain a product of the form $\Pi_{i=1}^{a_i} f(x_i)(1+O(S^{-\frac{1}{2}}))$, which does not yield divergence in the limit. To harness this strategy, we use a result that is analogous to Theorem 9 but starts with an approximation of $O(S^{-\frac{1}{2}-w\tau})$. For this additional result, stated in Theorem 10, we no longer need the full assumptions made in Theorem 9, since starting from an improved approximation means that certain terms must stay bounded.

Theorem 10. *Consider an approximation*

$$\frac{\|G_{\mathbf{d_{-i}}}\|}{\|G_{\mathbf{d_{-j}}}\|} = f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}, \sigma)(1 + O(S^{-\frac{1}{2}-w\tau}))$$

for some $w > 0$. Furthermore suppose that for $m = O(S^{\frac{1}{2}-\tau})$,

$$f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}_0, \sigma) = f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}_1, \sigma) + z(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}_0 - \mathbf{d}_1, \mathbf{d}_0, \sigma)$$

where $\|\mathbf{d}_1 - \mathbf{d}_0\|_1 \leq m$ and $z(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}_0 - \mathbf{d}_1, \mathbf{d}_0, \sigma) \leq O(S^{-\frac{1}{2}-\tau}) f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}_0, \sigma)$. If $\sigma = \mathbf{a}$, then we can construct a sharper approximation

$$\frac{\|G_{\mathbf{d_{-i}}}\|}{\|G_{\mathbf{d_{-j}}}\|} = g(\mathbf{e}_i, \mathbf{e_j}, \mathbf{d}, \mathbf{a})(1 + O(S^{-\frac{1}{2}-(w+2)\tau}))$$

where

$$\begin{aligned} g(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}, \mathbf{a}) \quad = \quad & \tfrac{a_i}{a_j} \exp(\log(1 + \frac{(a_i-1) \sum_{x_1 \neq \ldots \neq x_{a_{i-1}} = x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, (\mathbf{a} - a_i \mathbf{e}_i, (\mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} + \sum_{j=1}^{a_i-k} \mathbf{e}_{u_j}), \mathbf{b})}{\sum_{x_1 \neq \ldots \neq x_{a_{i-1}} \neq x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, (\mathbf{a} - a_i \mathbf{e}_i, \mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} + \sum_{j=1}^{a_i-k} \mathbf{e}_{u_j}), \mathbf{b})}) - \\ & \log(1 + \frac{(a_j-1) \sum_{x_1 \neq \ldots \neq x_{a_{j-1}} = x_{a_j}} \Pi_{k=1}^{a_j} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, (\mathbf{a} - a_j e_j, \mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} + \sum_{j=1}^{a_i-k} \mathbf{e}_{u_j}), \mathbf{b})}{\sum_{x_1 \neq \ldots \neq x_{a_{j-1}} \neq x_{a_j}} \Pi_{k=1}^{a_j} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, (\mathbf{a} - a_j e_j, \mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} + \sum_{j=1}^{a_i-k} \mathbf{e}_{u_j}), \mathbf{b})})) \end{aligned}$$

for an arbitrary choice of u_k. A similar result holds, with g depending on $\mathbf{b}$, if $\sigma = \mathbf{b}$.

The proof is very similar to the prior theorem so we leave the details to Appendix C. So in order to construct asymptotics for the ratio of the number of graphs of two slightly different degree sequences, starting with our approximation $\frac{\|G_{\mathbf{d}_{-\mathbf{i}}}\|}{\|G_{\mathbf{d}_{-\mathbf{j}}}\|} = \frac{a_i}{a_j}(1 + O(S^{-2\tau}))$, we apply Theorem 9, obtaining stronger approximations until we reach an approximation with a multiplicative error of $(1 + O(S^{-\frac{1}{2}}))$ from the solution. We can then apply Theorem 10 to construct approximations that are arbitrarily accurate. As mentioned before, this argument requires that the additional assumptions of Theorem 9 are true, which we will prove in Section 6.

We now provide a general method for constructing asymptotics for $\|G_{\mathbf{d}}\|$ from asymptotics for $\|G_{\mathbf{d}}\|/\|G_{\mathbf{d}_*}\|$. We start our derivation, as we did in Section 2, by considering a special case where it is particularly easy to count the number of graphs that realize a specified degree sequence.

Corollary 11. *Consider the degree sequence* $(\mathbf{a}, \mathbf{b}) \in \mathbb{Z}^{N\times 2}$ *where* 1 *appears exactly* S *times in* $\mathbf{b}$ *and* 0 *appears exactly* $N - S$ *times in* $\mathbf{b}$*. Then*

$$\|G_{(\mathbf{a},\mathbf{b})}\| = \frac{S!}{\Pi_{i=1}^{N} a_i!}$$

Proof. The result follows immediately from Lemma 1 where $q = 0$ and $k = N$. □

Theorem 11. *Define a function* $\rho : \mathbb{N} \to \mathbb{N}$ *such that* $\mathbf{b} = \sum_{k=1}^{S} \mathbf{e}_{\rho(k)}$*. Suppose that for each* $i \in \{1, \ldots, S\}$*, the following approximation holds:*

$$\frac{\|G_{(\mathbf{a},\sum_{k=1}^{S-i}\mathbf{e}_k+\sum_{k=1}^{i} e_{\rho(k)})}\|}{\|G_{(\mathbf{a},\sum_{k=1}^{S-i+1}\mathbf{e}_k+\sum_{k=1}^{i-1}\mathbf{e}_{\rho(k)})}\|}(1 + O(S^{-1-\epsilon})) = f(i) \tag{3.29}$$

for some $\epsilon > 0$*. Then*

$$\|G_{\mathbf{d}}\| = (1 + O(S^{-\epsilon}))\frac{S!}{\Pi_{i=1}^{N} a_i!}\Pi_{i=1}^{S} f(i).$$

Proof. For simplicity, we first suppose that $S \le N$. Write $\mathbf{1}_S = \sum_{k=1}^{S} \mathbf{e}_k$ where $\mathbf{e}_k$ is the kth standard unit vector. From Corollary 11,

$$\|G_{(\mathbf{a},\mathbf{1}_S)}\| = \frac{S!}{\Pi_{i=1}^{N} a_i!}. \tag{3.30}$$

Consequently,

$$\|G_{\mathbf{d}}\| = \|G_{(\mathbf{a},\mathbf{1}_S)}\|\Pi_{i=1}^{S}\frac{\|G_{(\mathbf{a},\sum_{k=1}^{S-i}\mathbf{e}_k+\sum_{k=1}^{i}\mathbf{e}_{\rho(k)})}\|}{\|G_{(\mathbf{a},\sum_{k=1}^{S-i+1}\mathbf{e}_k+\sum_{k=1}^{i-1}\mathbf{e}_{\rho(k)})}\|}. \tag{3.31}$$

But by assumption (3.29) and equation (3.30), we have that

$$\|G_{(\mathbf{a},\mathbf{1_S})}\|\Pi_{i=1}^{S}\frac{\|G_{(\mathbf{a},\sum_{k=1}^{S-i}\mathbf{e}_k+\sum_{k=1}^{i}\mathbf{e}_{\rho(k)})}\|}{\|G_{(\mathbf{a},\sum_{k=1}^{S-i+1}\mathbf{e}_k+\sum_{k=1}^{i-1}\mathbf{e}_{\rho(k)})}\|} = \frac{S!}{\Pi_{i=1}^{N}a_i!}\Pi_{i=1}^{S}f(i)(1+O(S^{-1-\epsilon})) = \tag{3.32}$$

$$(1+O(S^{-\epsilon}))\frac{S!}{\Pi_{i=1}^{N}a_i!}\Pi_{i=1}^{S}f(i). \tag{3.33}$$

Finally, in the event that $S > N$,we extend $\mathbf{d} \in \mathbb{Z}^{N\times 2}$ to a bidegree sequence $\mathbf{d}_* \in \mathbb{Z}^{S\times 2}$ by appending zeros to $\mathbf{d}$. The proof then proceeds analogously.

□

In the following section we will apply Theorems 9, 10 and 11 to construct asymptotics for the number of graphs of a given degree sequence. The cases worked out in the following section will not only provide extra intution about the veracity of the assumptions of Theorem 9, but will also lay the groundwork for the rigorous proof that the assumptions are indeed true.

3.4 SOME EXAMPLES ILLUSTRATING THE MAIN RESULT

We start with a technical lemma that will help us evaluate summations of the form $\sum_{x_1\neq\ldots\neq x_r}^{N}\Pi_{i=1}^{r}f(x_i)$. where the notation $\sum_{x_1\neq\ldots\neq x_r}^{N}$ tells us that for each $i \in \{1,2,..,r\}$, we take $x_i \in \{1,2,...,N\}$ and we are summing over all $\binom{N}{r}$ choices of r distinct elements in $\{1,2,...,N\}$.

Lemma 3. *Suppose that $f : \mathbb{N} \to [0,\infty)$, $\sum_{i=1}^{N} f(i) = S$, $\max f(\cdot) = O(S^{\frac{1}{2}-\tau})$, and that $x_1,\ldots,x_r$ is a sequence of natural numbers with $r = O(S^{\frac{1}{2}-\tau})$. Let k be a fixed (i.e., $O(1)$) natural number and define*

$$\xi = \sum_{\substack{x_1,..,x_k,\\ x_{k+1}\neq\ldots\neq x_r}}^{N}\Pi_{i=1}^{r}f(x_i).$$

Then

$$\sum_{x_1 \neq \ldots \neq x_r}^{N} \Pi_{i=1}^{r} f(x_i) = \xi(1 + O(S^{-2\tau})).$$

Proof. We proceed by induction on k. Start with $k = 1$. So

$$\sum_{x_1 \neq \ldots \neq x_r}^{N} \Pi_{i=1}^{r} f(x_i) = \sum_{x_1, x_2 \neq \ldots \neq x_r}^{N} \Pi_{i=1}^{r} f(x_i) - (r-1) \sum_{x_1 = x_2 \neq \ldots \neq x_r}^{N} \Pi_{i=1}^{r} f(x_2). \tag{3.34}$$

But the bounds on r and $\max f(\cdot)$ imply that

$$(r-1) \sum_{x_1 = x_2 \neq \ldots \neq x_r}^{N} \Pi_{i=1}^{r} f(x_i) \leq O(S^{1-2\tau}) \sum_{x_2 \neq \ldots \neq x_r}^{N} \Pi_{i=2}^{r} f(x_i), \tag{3.35}$$

while

$$\sum_{x_1, x_2 \neq \ldots \neq x_r}^{N} \Pi_{i=1}^{r} f(x_i) \geq O(S) \sum_{x_2 \neq \ldots \neq x_r}^{N} \Pi_{i=1}^{r} f(x_i). \tag{3.36}$$

Using (3.35) and (3.36) to express the right hand side of (3.34) in terms of ξ yields the desired result for $k = 1$.

Now assume that the lemma is true when $k = m$ and consider the case $k = m + 1$. We have

$$\sum_{x_1 \neq \ldots \neq x_r}^{N} \Pi_{i=1}^{r} f(x_i) = \sum_{\substack{x_1, .., x_m, \\ x_{m+1} \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i) + O(S^{-2\tau}) =$$

$$\sum_{\substack{x_1, .., x_m, x_{m+1}, \\ x_{m+2} \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i) + O(S^{-2\tau}) - (r - m - 1) \sum_{\substack{x_1, \ldots, x_m, \\ x_{m+1} = x_{m+2} \\ \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i).$$

By applying the estimates (3.35), (3.36), it follows as before that

$$(r - m - 1) \sum_{\substack{x_1, \ldots, x_m, \\ x_{m+1} = x_{m+2} \\ \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i) = O(S^{-2\tau}),$$

and the proof is complete.

□

Corollary 12. *With the notation of Lemma 3, let* $k > m > 1$ *be fixed natural numbers, let* $f, g : \mathbb{N} \to [0, \infty)$ *with* $g(\cdot) \leq f(\cdot)$, $\sum_{i=1}^{N} f(i) = S$, *and* $\max f(\cdot) = O(S^{\frac{1}{2}-\tau})$, *and let* $r = O(S^{\frac{1}{2}-\tau})$. *Define*

$$\zeta = \sum_{\substack{x_1=\ldots=x_m,\\ x_{m+1},\ldots,x_k,\\ x_{k+1}\neq\ldots\neq x_r}}^{N} g(x_1)\Pi_{i=2}^{r} f(x_i).$$

Then

$$\sum_{x_1=\ldots=x_m\neq\ldots\neq x_r}^{N} g(x_1)\Pi_{i=2}^{r} f(x_i) = \zeta(1 + O(S^{-2\tau})).$$

Proof. The proof is identical to that of Lemma 3. □

For use in later proofs, it is worth noting that while the error terms in Lemma 3 (and analogously in Corollary 12) are expressed as

$$O(S^{-2\tau}) \sum_{\substack{x_1,..,x_k,\\ x_{k+1}\neq\ldots\neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i),$$

they can also be stated in terms of $O(S^{-2\tau}) \sum_{x_1\neq\ldots\neq x_r}^{N} \Pi_{i=1}^{r} f(x_i)$ since asymptotically (as the above results imply),

$$\sum_{x_1\neq\ldots\neq x_r}^{N} \Pi_{i=1}^{r} f(x_i) \approx \sum_{\substack{x_1,..,x_k,\\ x_{k+1}\neq\ldots\neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i).$$

With Lemma 3 and Corollary 12 at our disposal, we can now derive the first order approximation of our power series with relative ease, but first we introduce some additional notation.

Definition 4. *Given a bidegree sequence* $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{Z}^{N\times 2}$, *we denote* $\alpha_k := \sum_{i=1}^{N} a_i^k$ *and* $\beta_k := \sum_{i=1}^{N} b_i^k$.

Theorem 12. *If* $\max_i \max(a_i, b_i) = O(S^{\frac{1}{2}-\tau})$ *for* $\tau > 0$, *then*

$$\frac{\|G_{d_{-i}}\|}{\|G_{d_{-j}}\|} = \frac{a_i}{a_j} e^{(a_i-a_j)\epsilon + O(S^{-4\tau})}$$

where, $\epsilon = (\beta_2 - \beta_1)/\beta_1^2$ *and* $(a_i - a_j)\epsilon = O(S^{-2\tau})$.

Proof. We start with the approximation

$$\frac{\|G_{\mathbf{d}_{-i}}\|}{\|G_{\mathbf{d}_{-j}}\|} = f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}, \mathbf{a})(1 + O(S^{-2\tau}))$$

for $f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}, \mathbf{a}) = a_i/a_j$. We can apply Theorem 9, since our approximation depends only on the degrees of the nodes i and j and we can decompose our approximation, $f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}, \mathbf{a}) = h(\mathbf{e}_i, \mathbf{d}, \mathbf{a})/h(\mathbf{e}_j, \mathbf{d}, \mathbf{a})$, where $h(\mathbf{e}_i, \mathbf{d}, \mathbf{a}) = a_i$;below, we shall also use $f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}, \mathbf{b}) = h(\mathbf{e}_i, \mathbf{d}, \mathbf{b})/h(\mathbf{e}_j, \mathbf{d}, \mathbf{b})$, where $h(\mathbf{e}_i, \mathbf{d}, \mathbf{b}) = b_i$ For the remainder of the proof, we omit the dependence on $\mathbf{d}$ in our notation.

From the conclusion of Theorem 9, we need to evaluate

$$\delta_i := \frac{(a_i - 1)\sum_{x_1 \neq \ldots \neq x_{a_i-1} = x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, \mathbf{b}_k)}{\sum_{x_1 \neq \ldots \neq x_{a_i-1} \neq x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, \mathbf{b}_k)}$$

and the analogously defined δ_j, where

$$\mathbf{b}_k = \mathbf{b} - \sum_{j=1}^{a_i-k+1} \mathbf{e}_{u_j} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j}$$

denotes the out-degree of the sequence $d_{k,i,j}$, as in the statement of Theorem 9. Now, multiplying our numerator and denominator by $\Pi_{k=1}^{a_i} b_{u_k}$ removes the dependence of f on $\mathbf{e}_{u_k}$, as $f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, \mathbf{b}_k) = b_{x_k}/b_{u_k}$ except for when $x_{a_i-1} = x_{a_i}$ and $k = a_i$, in which case $f(\mathbf{e}_{x_{a_i}}, \mathbf{e}_{u_{a_i}}, \mathbf{b}_{a_i}) = (b_{x_{a_i}} - 1)/b_{u_{a_i}}$, since we have already subtracted an outgoing edge from the node $b_{x_{a_i-1}}$, reducing its out-degree by 1.

By Corollary 12, with $g = f(\mathbf{e}_{x_{a_i}}, \mathbf{e}_{u_{a_i}}, \mathbf{b}_{a_i})$ and a relabeling of indices such that the case where $x_{a_i-1} = x_{a_i}$ becomes the case where $x_1 = x_2$, we have

$$\delta_i = \frac{(a_i - 1)\sum_{x_1 = x_2, x_3 \neq \ldots \neq x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{b}_k)}{\sum_{x_1, x_2, x_3 \neq \ldots \neq x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{b}_k)}(1 + O(S^{-4\tau})).$$

Factoring out $\sum_{x_3 \neq \ldots \neq x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{b})$ yields the following simplification:

$$\delta_i = \frac{(a_i - 1)\sum_{x_1 = x_2} \Pi f(\mathbf{e}_{x_k}, \mathbf{b}_k)}{\sum_{x_1, x_2} \Pi f(\mathbf{e}_{x_k}, \mathbf{b}_k)}(1 + O(S^{-4\tau})) = \frac{(a_i - 1)(\beta_2 - \beta_1)}{\beta_1^2}(1 + O(S^{-4\tau})).$$

where the last equality above follows from the fact that in the numerator where $x_1 = x_2$, $f(\mathbf{e}_{x_1}, \mathbf{b}_1) * f(\mathbf{e}_{x_2}, \mathbf{b}_2) = f(\mathbf{e}_{x_1}, \mathbf{b}_1) * f(\mathbf{e}_{x_1}, \mathbf{b}_2) = b_{x_1}(b_{x_1} - 1) = b_{x_1}^2 - b_{x_1}$. Summing over all choices for x_1 yields the expression $\beta_2 - \beta_1$.

We conclude that $\log(1+\delta_i) = \log(1+\frac{(a_i-1)[\beta_2-\beta_1]}{\beta_1^2}) + O(S^{-4\tau})$ and analogously $\log(1+\delta_j) = \log(1+\frac{(a_j-1)[\beta_2-\beta_1]}{\beta_1^2}) + O(S^{-4\tau})$. Hence from Theorem 9 we know that $\frac{\|G_{d\_i}\|}{\|G_{d\_j}\|} = \frac{a_i}{a_j}\exp[\log(1+\frac{(a_i-1)[\beta_2-\beta_1]}{\beta_1^2}) - \log(1+\frac{(a_j-1)[\beta_2-\beta_1]}{\beta_1^2}) + O(S^{-4\tau})]$ where writing the logarithms as a Taylor series yields $\frac{\|G_{d\_i}\|}{\|G_{d\_j}\|} = \frac{a_i}{a_j}\exp[\frac{(a_i-a_j)(\beta_2-\beta_1)}{\beta_1^2} + O(S^{-4\tau})]$.

Finally, note that

$$(a_i - a_j)\frac{\beta_2 - \beta_1}{\beta_1^2} \leq \frac{d_{max}\beta_2}{\beta_1^2} \leq \frac{d_{max}^2\beta_1}{\beta_1^2} = \frac{d_{max}^2}{\beta_1} = O(S^{-2\tau})$$

□

We can also prove the following statement with relative ease.

Theorem 13. *Using the notation from Definition 4 (with* $\alpha_1 = S$*), if* $\max_i \max(a_i, b_i) = O(S^{\frac{1}{4}-\epsilon})$ *for* $\epsilon > 0$*, then*

$$\|G_{\mathbf{d}}\| = [1 + O(S^{-4\epsilon})]\frac{S!}{\Pi_{i=1}^N a_i! b_i!}\exp(\frac{-(\alpha_2 - \alpha_1)(\beta_2 - \beta_1)}{2S^2}).$$

Proof. We know from Theorem 12 that

$$\frac{\|G_{d_{-i}}\|}{\|G_{d_{-j}}\|} = \frac{a_i}{a_j}e^{(a_i - a_j)\epsilon + O(S^{-4\tau})}$$

where $\epsilon = (\beta_2 - \beta_1)/\beta_1^2, (a_i - a_j)\epsilon = O(S^{-2\tau})$ when $d_{max} = O(S^{\frac{1}{2}-\tau})$ for $\tau \in (0, 1/2)$. If, in fact, $d_{max} = \max_i \max(a_i, b_i) = O(S^{\frac{1}{4}-\epsilon})$, then we have that $\tau = \frac{1}{4} + \epsilon$ and consequently $O(S^{-4\tau}) = O(S^{-1-4\epsilon})$. By Theorem 11 and the definition of ρ given in the theorem statement, we know that

$$\|G_{\mathbf{d}}\| = \frac{S!}{\Pi_{i=1}^N a_i!}\Pi_{i=1}^S f(i)(1 + O(S^{-4\epsilon}))$$

where by equation (3.29) we can approximate

$$f(i) = \frac{\|G_{(\mathbf{a}, \sum_{k=1}^{S-i} e_k + \sum_{k=1}^{i} e_{\rho(k)})}\|}{\|G_{(\mathbf{a}, \sum_{k=1}^{S-i+1} e_k + \sum_{k=1}^{i-1} e_{\rho(k)})}\|}(1 + O(S^{-1-4\epsilon})).$$

Now, we will proceed with a sequence of unit out-degree switches that transform an out-degree sequence consisting entirely of 1's to the out-degree sequence of $\mathbf{d}$, namely $\mathbf{b}$. Without loss of generality, suppose that the first node in $\mathbf{b}$ has out-degree b_1, such that there are $b_1 - 1$ switches required , and that $\rho(1) = \cdots = \rho(b_1 - 1) = 1$, which means that all of the switches will be applied to the first node, taking it from out-degree 1 to out-degree b_1. Then it follows from Theorem 12 that $f(1) = \frac{1}{2}\exp((\alpha_2 - \alpha_1)(1-2)/S^2)$, since the first switch changes the out-degree of the first node from 1 to 2. Similarly, it follows that $f(2) = \frac{1}{3}\exp((\alpha_2 - \alpha_1)(1-3)/S^2)$ and, more generally for $k \leq b_1 - 1$, $f(k) = \frac{1}{k+1}\exp((\alpha_2 - \alpha_1)(1-[k+1])/S^2) = \frac{1}{k+1}\exp((\alpha_2-\alpha_1)(-k)/S^2)$. Hence

$$\Pi_{k=1}^{b_i-1} f(k) = \Pi_{k=1}^{b_i-1}\frac{1}{k+1}\exp((\alpha_2-\alpha_1)(-k)/S^2) =$$

$$\frac{\exp((\alpha_2-\alpha_1)\sum_{k=1}^{b_1-1}(-k)/S^2)}{b_1!} = \frac{\exp(-\frac{(\alpha_2-\alpha_1)(b_1^2-b_1)}{2S^2})}{b_1!}.$$

Repeating this argument for all of the nodes in the degree sequence yields the result,

$$\|G_{\mathbf{d}}\| = (1+O(S^{-4\epsilon}))\frac{S!}{\Pi_{i=1}^N a_i!}\Pi_{i=1}^S f(i) = (1+O(S^{-4\epsilon}))\frac{S!}{\Pi_{i=1}^N a_i! b_i!}\exp(\frac{-(\alpha_2-\alpha_1)(\beta_2-\beta_1)}{2S^2}).$$

□

We now explain the intuition behind some generalizations of Lemma 3 that we need to achieve higher order approximations. Note that in Theorem 12, using Lemma 3, we performed an approximation of the form

$$\frac{\sum_{x_1=x_2\neq\ldots\neq x_r}\Pi_{n=1}^r f(x_n)}{\sum_{x_1\neq x_2\neq\ldots\neq x_r}\Pi_{n=1}^r f(x_n)} \approx \frac{\sum_{x_1=x_2}\Pi_{n=1}^2 f(x_n)}{\sum_{x_1,x_2}\Pi_{n=1}^2 f(x_n)}\left[\frac{\sum_{x_3\neq\ldots\neq x_r}\Pi_{n=3}^r f(x_n)}{\sum_{x_3\neq\ldots\neq x_r}\Pi_{n=3}^r f(x_n)}\right]$$

$$= \frac{\sum_{x_1=x_2}\Pi_{n=1}^2 f(x_n)}{\sum_{x_1,x_2}\Pi_{n=1}^2 f(x_n)}$$

and we showed that such an approximation yielded an $O(S^{-4\tau})$ error.

More generally, we want to construct approximations of

$$\sum_{x_1=x_2\neq\ldots\neq x_r}^{N}\Pi_{n=1}^r f(x_n) \text{ and } \sum_{x_1\neq x_2\neq\ldots\neq x_r}^{N}\Pi_{n=1}^r f(x_n)$$

For example, to attain a generalization of Lemma 2, we allow for the possibility (in the numerator) that $x_1 = x_2 = x_3$, but now we need to separate out three terms as opposed to two to achieve our desired cancellation; that is, we have

$$\frac{\sum_{x_1=x_2=x_3\neq\ldots\neq x_r}\Pi_{n=1}^r f(x_n)}{\sum_{x_1\neq x_2\neq\ldots\neq x_r}\Pi_{n=1}^r f(x_n)} \approx \frac{\sum_{x_1=x_2=x_3}\Pi_{n=1}^3 f(x_n)}{\sum_{x_1,x_2,x_3}\Pi_{n=1}^3 f(x_n)}\left[\frac{\sum_{x_4\neq\ldots\neq x_r}\Pi_{n=3}^r f(x_n)}{\sum_{x_4\neq\ldots\neq x_r}\Pi_{n=3}^r f(x_n)}\right].$$

However, once we are computing $O(S^{-4\tau})$ terms, we also need to consider the case where $x_1 = x_2$ and $x_3 = x_4$, as such terms also turn out to contribute at $O(S^{-4\tau})$. To make this idea more rigorous, consider approximating

$$\begin{aligned}&\textstyle\sum_{x_1=x_2\neq x_3\neq\ldots\neq x_r}\Pi_{n=1}^r f(x_n) = \\ &\textstyle\sum_{x_1=x_2,x_3\neq\ldots\neq x_r}\Pi_{n=1}^r f(x_n) - (r-2)\sum_{x_1=x_2=x_3\neq\ldots\neq x_r}\Pi_{n=1}^r f(x_n).\end{aligned}$$

To obtain a sufficiently high order estimate for the left hand side, we partition the right hand side into still more terms, motivated by the intuition that terms with more equal signs under a summation should be of higher order; thus, we consider

$$\sum_{x_1=x_2\neq x_3\neq\ldots\neq x_r}\Pi_{n=1}^r f(x_n) = \sum_{\substack{x_1=x_2,x_3,\\ x_4\neq\ldots\neq x_r}}\Pi_{n=1}^r f(x_n) - (r-3)\sum_{\substack{x_1=x_2,\\ x_3=x_4\neq\ldots\neq x_r}}\Pi_{n=1}^r f(x_n) -$$

$$(r-2)\sum_{\substack{x_1=x_2=x_3,\\ x_4\neq\ldots\neq x_r}}\Pi_{n=1}^r f(x_n) + \ldots$$

where we have neglected to write out the final terms of still higher order.

Now, since we want to keep the case where $x_1 = x_2$ and $x_3 = x_4$, we have to integrate out the x_4, which yields

$$\sum_{x_1=x_2\neq x_3\neq\ldots\neq x_r}\Pi_{n=1}^r f(x_n) = \sum_{\substack{x_1=x_2,\\ x_3,x_4,x_5\neq\ldots\neq x_r}}\Pi_{n=1}^r f(x_n) - (r-3)\sum_{\substack{x_1=x_2,x_3=x_4,\\ x_5\neq\ldots\neq x_r}}\Pi_{n=1}^r f(x_n)$$

$$-(r-2)\sum_{\substack{x_1=x_2=x_3,x_4,\\ x_5\neq\ldots\neq x_r}}\Pi_{n=1}^r f(x_n) - (r-4)\sum_{\substack{x_1=x_2,x_3,\\ x_4=x_5\neq\ldots\neq x_r}}\Pi_{n=1}^r f(x_n) + \ldots$$

But we can simply relabel indices to observe that

$$\sum_{x_1=x_2,x_3,x_4=x_5\neq\ldots\neq x_r}\Pi_{n=1}^r f(x_n) = \sum_{\substack{x_1=x_2,x_3=x_4,\\ x_5\neq\ldots\neq x_r}}\Pi_{n=1}^r f(x_n) + \ldots,$$

and we conclude that

$$\sum_{x_1=x_2\neq x_3\neq \ldots \neq x_r} \Pi_{n=1}^r f(x_n) = \sum_{x_1=x_2,x_3,x_4,x_5\neq\ldots\neq x_r} \Pi_{n=1}^r f(x_n)$$

$$-(2r-7)\sum_{\substack{x_1=x_2,x_3=x_4,\\ x_5\neq\ldots\neq x_r}} \Pi_{n=1}^r f(x_n) - (r-2)\sum_{\substack{x_1=x_2=x_3,x_4,\\ x_5\neq\ldots\neq x_r}} \Pi_{n=1}^r f(x_n) + \ldots$$

So in fact, it suffices to factor out four indices. Indeed, the general intuition is that if we are including k equalities in our expansion, then we will need to factor out $2k$ indices, to consider the worst case scenario where $x_1 = x_2, x_3 = x_4, ..., x_{2k-1} = x_{2k}$. But we can then relabel indices to express terms with, say, $x_1 = x_2, x_3, x_4 = x_5, ..., x_{2k} = x_{2k+1}$ using terms of the form $x_1 = x_2, x_3 = x_4, ..., x_{2k-1} = x_{2k}$.

With these ideas in mind, we now proceed to derive the higher order approximation. We subdivide this task into a few steps, starting with an approximation lemma on sums of the form $\sum_{x_1\neq\ldots\neq x_r} \Pi_{n=1}^r f(x_n)$ and sums of the form $\sum_{x_1=x_2\neq\ldots\neq x_r} \Pi_{n=1}^r f(x_n)$.

Lemma 4. *Let $f, g : \mathbb{N} \to [0,\infty)$ with $f(\cdot) \geq g(\cdot)$, $\sum_{i=1}^N f(i) = S$, $\max f(\cdot) = O(S^{\frac{1}{2}-\tau})$, and $r = O(S^{\frac{1}{2}-\tau})$, $r \geq 4$ and define $F = \Pi_{i=1}^4 f(x_i)$, $G = g(x_1)\Pi_{i=2}^4 f(x_i)$,*

$$\tilde{\xi} = \sum_{\substack{x_1,\ldots,x_4,\\ x_5\neq\ldots\neq x_r}}^N \Pi_{i=1}^r f(x_i) - (4r-10)\sum_{\substack{x_1=x_2,x_3,x_4,\\ x_5\neq\ldots\neq x_r}}^N \Pi_{i=1}^r f(x_i) \quad \textit{and}$$

$$\begin{aligned}\tilde{\zeta} = \sum_{\substack{x_1=x_2,x_3,x_4,\\ x_5\neq\ldots\neq x_r}}^N g(x_1)\Pi_{i=2}^r f(x_i) \quad &- \quad (r-2)\sum_{\substack{x_1=x_2=x_3,x_4,\\ x_5\neq\ldots\neq x_r}}^N g(x_1)\Pi_{i=2}^r f(x_i) \\ &- \quad (2r-7)\sum_{\substack{x_1=x_2,x_3=x_4,\\ x_5\neq\ldots\neq x_r}}^N g(x_1)\Pi_{i=2}^r f(x_i).\end{aligned}$$

Then

$$\xi_0 := \sum_{x_1\neq\ldots\neq x_r}^N \Pi_{i=1}^r f(x_i) = \tilde{\xi}(1 + O(S^{-4\tau})) \tag{3.37}$$

and

$$\zeta_0 := \sum_{x_1=x_2\neq\ldots\neq x_r} g(x_1)\Pi_{i=2}^r f(x_i) = \tilde{\zeta}(1 + O(S^{-4\tau})). \tag{3.38}$$

Furthermore,

$$\frac{\zeta_0}{\xi_0} = \frac{\sum_{x_1=x_2,x_3,x_4} G - (r-2)\sum_{x_1=x_2=x_3,x_4} G - (2r-7)\sum_{x_1=x_2,x_3=x_4} G}{\sum_{x_1,..,x_4} F - (4r-10)\sum_{x_1=x_2,x_3,x_4} F} + O(S^{-\frac{1}{2}-5\tau}). \tag{3.39}$$

Proof. To derive the first equality, start by expressing ξ_0 as

$$\xi_0 = \sum_{x_1 \neq \ldots \neq x_r}^{N} \Pi_{i=1}^{r} f(x_i) = \sum_{\substack{x_1, \\ x_2 \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i) - (r-1) \sum_{\substack{x_1 = x_2 \neq \\ x_3 \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i).$$

Applying Lemma 2 to the final term yields

$$\xi_0 = \sum_{\substack{x_1, \\ x_2 \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i) - (r-1) \sum_{\substack{x_1 = x_2, x_3, x_4, \\ x_5 \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i) + O(S^{-4\tau}) \sum_{x_1 \neq \ldots \neq x_r}^{N} \Pi_{i=1}^{r} f(x_i)$$

where the error term comes from

$$O(S^{-2\tau})(r-1) \sum_{\substack{x_1 = x_2, x_3, x_4 \\ x_5 \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i) = O(S^{-4\tau}) \sum_{x_1 \neq \ldots \neq x_r}^{N} \Pi_{i=1}^{r} f(x_i). \tag{3.40}$$

We next repeat the same argument and relabel to obtain

$$\begin{aligned} \xi_0 &= \textstyle\sum_{\substack{x_1, x_2, \\ x_3 \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i) - (r-1) \sum_{\substack{x_1 = x_2, x_3, x_4, \\ x_5 \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i) - \\ &(r-2) \textstyle\sum_{\substack{x_1, x_2 = x_3, x_4, \\ x_5 \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i) + O(S^{-4\tau}) \sum_{x_1 \neq \ldots \neq x_r}^{N} \Pi_{i=1}^{r} f(x_i) \\ &= \textstyle\sum_{\substack{x_1, x_2, \\ x_3 \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i) - (2r-3) \sum_{\substack{x_1 = x_2, x_3, x_4, \\ x_5 \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i) + \\ &O(S^{-4\tau}) \textstyle\sum_{x_1 \neq \ldots \neq x_r}^{N} \Pi_{i=1}^{r} f(x_i). \end{aligned} \tag{3.41}$$

Continuing in this fashion,

$$\xi_0 = \sum_{\substack{x_1, x_2, x_3, \\ x_4 \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i) - (3r-6) \sum_{\substack{x_1 = x_2, x_3, x_4, \\ x_5 \neq \ldots \neq x_r}}^{N} \Pi_{i=1}^{r} f(x_i) + O(S^{-4\tau}) \sum_{x_1 \neq \ldots \neq x_r}^{N} \Pi_{i=1}^{r} f(x_i).$$

The next, final step is a bit trickier so we include a little more detail. First, we extract the x_4 out of the first summation of the above equation and apply (3.40), which yields

$$\xi_0 = \sum_{\substack{x_1,x_2,x_3 x_4,\\ x_5\neq\ldots\neq x_r}}^N \Pi_{i=1}^r f(x_i) - (3r-6)\sum_{\substack{x_1=x_2,x_3,x_4,\\ x_5\neq\ldots\neq x_r}}^N \Pi_{i=1}^r f(x_i)$$

$$-(r-4)\sum_{\substack{x_1,x_2,x_3,x_4=x_5,\\ x_6\neq\ldots\neq x_r}}^N \Pi_{i=1}^r f(x_i) + O(S^{-4\tau})\sum_{x_1\neq\ldots\neq x_r}^N \Pi_{i=1}^r f(x_i).$$

But by another application of (3.40),

$$\sum_{\substack{x_1,x_2,x_3,x_4=x_5,\\ x_6\neq\ldots\neq x_r}}^N \Pi_{i=1}^r f(x_i) = \sum_{\substack{x_2,x_3,x_4=x_5,\\ x_1\neq x_6\neq\ldots\neq x_r}}^N \Pi_{i=1}^r f(x_i) + O(S^{-4\tau})\sum_{x_1\neq\ldots\neq x_r}^N \Pi_{i=1}^r f(x_i),$$

where by relabeling, the first term on the right hand side is the same as

$$\sum_{\substack{x_1=x_2,x_3,x_4,\\ x_5\neq\ldots\neq x_r}}^N \Pi_{i=1}^r f(x_i).$$

So we conclude that

$$\xi_0 = \sum_{\substack{x_1,x_2,x_3,x_4,\\ x_5\neq\ldots\neq x_r}}^N \Pi_{i=1}^r f(x_i) - (4r-10)\sum_{\substack{x_1=x_2,x_3,x_4,\\ x_5\neq\ldots\neq x_r}}^N \Pi_{i=1}^r f(x_i) + O(S^{-4\tau})\sum_{x_1\neq\ldots\neq x_r}^N \Pi_{i=1}^r f(x_i),$$

where the sum in the final term is itself ξ_0, such that the first part of the Lemma is established.

The second part of the Lemma follows analogously, while expression (3.39) follows from cancellation of terms. □

Theorem 14. *Let $d_{max} = O(S^{\frac{1}{2}-\tau})$. Then*

$$\frac{\|G_{d_{-i}}\|}{\|G_{d_{-j}}\|} = \frac{a_i}{a_j}\exp([a_i - a_j]\epsilon_1 - [a_i^2 - a_j^2]\epsilon_2 + O(S^{\max(-6\tau,-\frac{1}{2}-3\tau)}))$$

where

$$\epsilon_1 = \frac{\beta_2 + 2\beta_3\alpha_2/\alpha_1^2}{(\beta_1 + \beta_2\alpha_2/\alpha_1^2)^2} \text{ and } \epsilon_2 = \frac{(\beta_2-\beta_1)^2}{2{\beta_1}^4} + \frac{\beta_3\beta_1 - 2\beta_2^2}{\beta_1^4}.$$

Proof. From Theorem 12 we know that

$$\frac{\|G_{\mathbf{d_{-i}}}\|}{\|G_{\mathbf{d_{-j}}}\|} = \frac{a_i}{a_j}\exp((a_i - a_j)\frac{\beta_2 - \beta_1}{\beta_1^2} = \frac{a_i \exp(a_i \frac{\beta_2}{\beta_1^2})}{a_j \exp(a_j \frac{\beta_2}{\beta_1^2})}(1 + O(S^{-4\tau})). \tag{3.42}$$

Note that the middle expression in (3.42) satisfies the assumption from Theorem 9 that our approximation can be expressed as $h(i, \mathbf{d})/h(j, \mathbf{d})$ for some function h, and hence we can use Theorem 9 to derive the next step on our expansion. Similarly if we consider a different degree sequence such that $\|\mathbf{d}_0 - \mathbf{d}_1\|_\infty = 1$, $\|\mathbf{d}_0\|_1 = \|\mathbf{d}_1\|_1$ and $\|\mathbf{d}_0 - \mathbf{d}_1\|_1 = O(S^{\frac{1}{2}-\tau})$, then we introduce a multiplicative error of $(1 + O(S^{-6\tau}))$ as

$$\frac{a_i}{a_j}\exp((a_i - a_j)\frac{\beta_2 - \beta_1}{\beta_1^2}$$

only depends on the second moment of the out-degree sequence β_2. We introduce the notation $\beta_2(\mathbf{d})$ to explicitly denote the dependence of β_2 on the degree sequence.

Consequently, the difference between $|\beta_2(\mathbf{d}_1) - \beta_2(\mathbf{d}_0)| = O(S^{1-2\tau})$, hence

$$\frac{\frac{a_i}{a_j}\exp((a_i - a_j)\frac{\beta_2(\mathbf{d_1})}{\beta_1^2})}{\frac{a_i}{a_j}\exp((a_i - a_j)\frac{\beta_2(\mathbf{d_0})}{\beta_1^2})} = \exp((a_i - a_j)\frac{\beta_2(\mathbf{d_1}) - \beta_2(\mathbf{d_0})}{\beta_1^2}) =$$

$$\exp(O(S^{\frac{1}{2}-\tau}O(S^{1-2\tau})/O(S^2)) = \exp(O(S^{-\frac{1}{2}-3\tau})).$$

We now proceed as we did in Theorem 12. We need to evaluate

$$\delta_i = \frac{(a_i - 1)\sum_{x_1 \neq \ldots \neq x_{a_{i-1}} = x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, \mathbf{b})}{\sum_{x_1 \neq \ldots \neq x_{a_{i-1}} \neq x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, \mathbf{b})}.$$

We multiply our numerator and denominator by $\Pi_{k=1}^{a_i} b_{u_k}\exp(b_{u_k}\frac{\alpha_2}{\alpha_1^2})$, where α_1, α_2 are the moments of the degree sequence $\mathbf{d}$, we drop the dependence of f on $\mathbf{e}_{u_k}$, and we define $f(\mathbf{e}_{x_k}, \mathbf{b}) = b_{x_k}\exp(b_{x_k}\frac{\alpha_2}{\alpha_1^2})$ to obtain (from (3.42)

$$\delta_i = \frac{(a_i - 1)\sum_{x_1 \neq \ldots \neq x_{a_{i-1}} = x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{b})}{\sum_{x_1 \neq \ldots \neq x_{a_{i-1}} \neq x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{b})}.$$

We can now invoke Lemma 14 to simplify the above summation. Further algebraic simplification yields the desired result.

□

To indicate how the estimates that we derive can, in theory, be continued indefinitely, we conclude this section by providing a higher order approximation for a ratio of the numbers of graphs that realize two different degree sequences that has an $O(S^{-8\tau})$ correction. As before, we need a more refined version of Lemma 3 to derive this approximation. We omit the proof, since the same proof technique that derived Lemma 3 yields the proof of Lemma 4. On the author's webpage, we provide code that computes (and proves) the refined variants of Lemma 3 up to arbitrary order.

Lemma 5. *Suppose that* $g : \mathbb{N} \to [0, \infty)$, $f : \mathbb{N} \to [1, \infty)$ *with* $f(\cdot) \geq g(\cdot)$, *that* $\sum_{i=1}^N f(i) = S$, $\max f(\cdot) = O(S^{\frac{1}{2}-\tau})$, *and that* $r = O(S^{\frac{1}{2}-\tau})$, $r \geq 6$. *Define* $\chi = \sum_{x_1 \neq \ldots \neq x_r}^N \Pi_{i=1}^r f(x_i)$ *and* $\kappa = \sum_{x_1 = x_2 \neq \ldots \neq x_r}^N g(x_1)\Pi_{i=2}^r f(x_i)$ *Furthermore, for notational convenience let* $F = \Pi_{i=1}^6 f(x_i)$ *and* $G = g(x_1)\Pi_{i=2}^6 f(x_i)$. *Then*

$$\frac{\kappa}{\chi} = \frac{\kappa_1}{\chi_1} + O(S^{-\frac{1}{2}-7\tau})$$

where

$$\chi_1 = \sum_{x_1,..,x_6} F - (6r-21) \sum_{\substack{x_1=x_2 \\ x_3..,x_6}} F + (9r^2 - 58r + 69) \sum_{\substack{x_1=x_2,x_3=x_4 \\ x_5,x_6}} F + (6r^2 - 48r + 112) \sum_{\substack{x_1=x_2=x_3 \\ x_4..,x_6}} F$$

and

$$\kappa_1 = \sum_{x_1=x_2,x_3..,x_6} G - (r-2) \sum_{\substack{x_1=x_2=x_3 \\ x_4..,x_6}} G - (4r-18) \sum_{\substack{x_1=x_2,x_3=x_4 \\ x_5,x_6}} G + (r^2 - 5r + 6) \sum_{\substack{x_1=x_2=x_3=x_4 \\ x_5,x_6}} G$$

$$+ (3r^2 - 21r + 30) \sum_{\substack{x_1=x_2=x_3 \\ x_4=x_5,x_6}} G + (4r^2 - 40r + 104) \sum_{\substack{x_1=x_2 \\ x_3=x_4=x_5,x_6}} G + (2r^2 - 15r + 21) \sum_{\substack{x_1=x_2 \\ x_3=x_4,x_5=x_6}} G.$$

We conclude this section with a statement of the corresponding more refined approximation.

Theorem 15. *Let $d_{max} = O(S^{\frac{1}{2}-\tau})$ and recall that $\beta_k = \sum_{k=1}^{N} b_i^k$. We have*

$$\frac{\|G_{d_{-i}}\|}{\|G_{d_{-j}}\|} = \frac{a_i}{a_j} \exp([a_i - a_j]\epsilon_1 - [a_i^2 - a_j^2]\epsilon_2 + [a_i^3 - a_j^3]\epsilon_3 + O(S^{\max(-1-2\tau,-8\tau)}))$$

where we define

$$\epsilon_1 = \frac{E_{\mathbf{b}}[f(x)f(x-1)]}{E_{\mathbf{b}}[f(x)]^2} + \frac{E_{\mathbf{b}}[f(x)f(x-1)]^2 - 5E_{\mathbf{b}}[f(x)^2]E_{\mathbf{b}}[f(x)f(x-1)] + 3E_{\mathbf{b}}[f(x)^2 f(x-1)]E_{\mathbf{b}}[f(x)]}{E_{\mathbf{b}}[f(x)]^4},$$

$$\epsilon_2 = \frac{-2E_{\mathbf{b}}[f(x)^2]E_{\mathbf{b}}[f(x)f(x-1)] + \frac{1}{2}E_{\mathbf{b}}[f(x)f(x-1)]^2 + E_{\mathbf{b}}[f(x)^2 f(x-1)]E_{\mathbf{b}}[f(x)]}{E_{\mathbf{b}}[f(x)]^4}, \quad \textit{and}$$

$$\epsilon_3 = \frac{-\frac{107}{3}\beta_2^3 - \frac{11}{2}\beta_1\beta_2\beta_3 + \beta_2\beta_4}{\beta_1^6}$$

for

$$E_{\mathbf{b}}[f(x)] := \textstyle\sum_{x \in \mathbf{b}} f(x) \ f(x) = x + x^2\eta_1 + \frac{x^3\eta_1^2}{2} - x^3\eta_2,$$

$$\eta_1 := \frac{\alpha_2 + 2\alpha_3\beta_2/\alpha_1^2}{(\alpha_1 + \alpha_2\beta_2/\alpha_1^2)^2}, \quad \textit{and} \ \ \eta_2 := \frac{(\alpha_2-\alpha_1)^2}{2{\alpha_1}^4} + \frac{\alpha_3\alpha_1 - 2\alpha_2^2}{\alpha_1^4}.$$

3.5 VERIFYING THE ASSUMPTIONS

To prove that the assumptions made in Theorem 9 (and in Theorem 10) hold, it is helpful to have results to simplify the analysis of ratios of sums of the form $\sum_{x_1 \neq \ldots \neq x_r} \Pi_{i=1}^r f(x_i)$ that are more general than the results we proved in Section 5. First, it is helpful to introduce some notation.

Definition 5. *Define the set of natural numbers $I = \{1, \ldots, r\}$. For any collection $A_=$ of disjoint subsets of I, we say that a sequence $\mathbf{x} \in \mathbb{N}^r$ satisfies $A_=$ if for every $\{i_1, \ldots, i_m\} \in A_=$, we have $x_{i_1} = x_{i_2} = \ldots = x_{i_m}$. Given that $\mathbf{x}$ satisfies $A_=$ and that $A_{\neq}$ is a subset of I, we say that $\mathbf{x}$ satisfies $A_{\neq}$ if for all $i, j \in A_{\neq}$ such that i, j are not both contained in any $A \in A_=$, we have $x_i \neq x_j$. We denote the set of all $\mathbf{x}$ that satisfy both $A_=$ and $A_{\neq}$ by the notation $A_= \otimes A_{\neq}$.*

Before proceeding with the theorem statements that enable us to construct approximations for ratios of sums of the form $\sum_{x_1\neq\ldots\neq x_r}\Pi_{i=1}^r f(x_i)$, we wish to motivate some of the notation that we use. To approximate $\sum_{x_1\neq\ldots\neq x_r}\Pi_{i=1}^r f(x_i)$, we wish to remove some (fixed) number of $x_i's$ from the list of required inequalities. For example, we can express $\sum_{x_1\neq x_2} f(x_1)f(x_2) = \sum_{x_1,x_2} f(x_1)f(x_2) - \sum_{x_1=x_2} f(x_1)f(x_2)$, where in the first term on the right hand side, we no longer have to worry about the 'does not equal' relationship of x_1 and x_2 found on the left hand side. To help us identify the dominating terms, we introduce sets of the form $A_{=}^{(i,c)}\otimes A_{\neq}^{(i,c)}$, where the i denotes the number of equal signs found under the summation and c denotes a particular instance where i of these variables are equal. For example, suppose we want to simplify $\sum_{x_1\neq x_2\neq x_3} f(x_1)f(x_2)f(x_3) = \sum_{x_1,x_2\neq x_3} f(x_1)f(x_2)f(x_3) - \sum_{x_1=x_2\neq x_3} f(x_1)f(x_2)f(x_3) - \sum_{x_1=x_3\neq x_2} f(x_1)f(x_2)f(x_3)$. On the right hand side there are two summations involving one equality under the summation. Hence we can define $A_{=}^{(1,0)}\otimes A_{\neq}^{(1,0)} = \{\{1,2\}\}\otimes\{1,2,3\}$, $A_{=}^{(1,1)}\otimes A_{\neq}^{(1,1)} = \{\{1,3\}\}\otimes\{1,2,3\}$ representing those two summations. Note that in each of these examples $i=1$, but since there are two possible choices, c can be either 0 or 1. Now we present two results regarding the sums of interest. Since the proofs of the two theorems are essentially identical, we only prove one of them.

Theorem 16. *Suppose that* $\sum_{x_m} f(x_m) = S$ *and that* $\max f(x_m) = O(S^{\frac{1}{2}-\tau})$ *and* $r = O(S^{\frac{1}{2}-\tau})$. *Consider* $\sum_{x_1\neq\ldots\neq x_r}\Pi_{m=1}^r f(x_m)$. *Fix a natural number* k *such that* $2k\leq r$ *(where* $2k$ *is the number of variables that we remove from the list of inequalities so that they can equal other variables). Then for all* $j\leq k$ *we can write*

$$\sum_{x_1\neq\ldots\neq x_r}\Pi_{m=1}^r f(x_m) = \sum_{i=0}^{j}\sum_{c=0}^{h(i,k)} p_{(i,c)} \sum_{A_{=}^{(i,c)}\otimes A_{\neq}^{(i,c)}} \Pi_{m=1}^r f(x_m) \tag{3.43}$$

where for each i, $h(i,k)+1$ *represents a number of arrangements of variables with* i *variables equal to each other, and* $h(0,k)=0$*; for all* i,c, $p_{(i,c)}$ *is a polynomial in* r*; and for all* $i<j$ *and for all* c, $A_{=}^{(i,c)}$ *consists of a collection of subsets of* $\{1,\ldots,2k\}$ *with* $\sum_{A\in A_{=}^{(i,c)}}(|A|-1) = i$ *and* $A_{\neq}^{(i,c)} = \{2k+1,...,r\}$. *In addition,*

- *for all* c, $\sum_{A\in A_{=}^{(j,c)}}(|A|-1) = j$,
- $A_{\neq}^{(j,c)} = \{s, s+1, ..., r\}$ *for some* $s\leq 3k$, *and*

- *if* $A \in A_{=}^{(j,c)}$ *has a nontrivial intersection with* $A_{\neq}^{(j,c)}$ *then* $A = \{s, s+1, ..., s+t\}$ *for some* t *such that* $s+t < r$*; moreover, there can only be one such* $A \in A_{=}^{(j,c)}$ *that has a nontrivial intersection with* $A_{\neq}^{(j,c)}$.

Finally,

$$\frac{p_{(i,c)} \sum_{A_{=}^{(i,c)} \otimes A_{\neq}^{(i,c)}} \Pi_{m=1}^{r} f(x_m)}{\sum_{x_1,...,x_{2k},x_{2k+1} \neq ... \neq x_r} \Pi_{m=1}^{r} f(x_m)} = O(S^{-2i\tau}).$$

Theorem 17. *Suppose that* $\sum_{x_m} f(x_m) = S$ *and that* $\max f(x_m) = O(S^{\frac{1}{2}-\tau})$ *and* $r = O(S^{\frac{1}{2}-\tau})$*. Consider* $\sum_{x_1=x_2\neq...\neq x_r} g(x_1)\Pi_{m=2}^{r} f(x_m)$*. For simplicity, we assume that the elements in* $A_{=}$ *are disjoint. Fix a number* k *(where* $2k$ *is the number of variables removed from the list of inequalities). Then for all* $j \leq k$ *we can write*

$$\sum_{x_1=x_2\neq...\neq x_r} g(x_1)\Pi_{m=2}^{r} f(x_m) = \sum_{i=0}^{j} \sum_{c=0}^{h(i,k)} p_{(i,c)} \sum_{A_{=}^{(i,c)} \otimes A_{\neq}^{(i,c)}} g(x_1)\Pi_{m=2}^{r} f(x_m)$$

where for each i*,* $h(i,k)+1$ *represents a number of arrangements of variables with* i *variables equal to each other, and* $h(0,k) = 0$*; for all* i, c*,* $p_{(i,c)}$ *is a polynomial in* r*; and for all* $i < j$ *and for all* c*,* $A_{=}^{(i,c)}$ *consists of a collection of subsets of* $\{1, ..., 2k\}$*,* $A_{\neq}^{(i,c)} = \{2k+1, ..., r\}$ *with* $\sum_{A \in A_{=}^{(i,c)}} (|A| - 1) = i + 1$*. In addition,*

- *for all* c*,* $\sum_{A \in A_{=}^{(j,c)}} (|A| - 1) = j + 1$*,*
- $A_{\neq}^{(j,c)} = \{s, s+1, ..., r\}$ *for some* $s \leq 3k$*, and*
- *if* $A \in A_{=}^{(j,c)}$ *has a nontrivial intersection with* $A_{\neq}^{(j,c)}$ *then* $A = \{s, s+1, ..., s+t\}$ *for some* t *such that* $s+t < r$*; moreover, there can only be one such* $A \in A_{=}^{(j,c)}$ *that has a nontrivial intersection with* $A_{\neq}^{(j,c)}$.

  Finally, for all $i \leq j$*,*

$$\frac{p_{(i,c)} \sum_{A_{=}^{(i,c)} \otimes A_{\neq}^{(i,c)}} g(x_1)\Pi_{m=2}^{r} f(x_m)}{\sum_{x_1=x_2,x_3,...,x_{2k},x_{2k+1} \neq ... \neq x_r} g(x_1)\Pi_{m=2}^{r} f(x_m)} = O(S^{-2i\tau}).$$

Proof of Theorem 16. We proceed by induction. For the base case of $j = 1$, we start by showing that the stated conditions hold in the case of $i = j$ (such that the bulleted list in the theorem statement applies) and then we come back to $(i, c) = (0, 0)$. We have seen previously that

$$\sum_{x_1 \neq \ldots \neq x_r} \Pi_{m=1}^r f(x_m) = \sum_{x_1, x_2 \neq \ldots \neq x_r} \Pi_{m=1}^r f(x_m) - (r-1) \sum_{x_1 = x_2 \neq \ldots \neq x_r} \Pi_{m=1}^r f(x_m).$$

We denote the set of $\mathbf{x}$ over which the final summation occurs as $A_{=}^{(1,0)} \otimes A_{\neq}^{(1,0)}$ where $A_{=}^{(1,0)} = \{\{1,2\}\}$, the variables that equal each other, and $A_{\neq}^{(1,0)} = \{1, 2, ..., r\}$, the variables that are required not to equal some (at least one) other variable. The coefficient $p_{(1,0)} = -(r-1)$ is also a polynomial in r and $\sum_{A \in A_{=}^{(1,0)}} (|A| - 1) = 1$ (where we use the superscript $(1, 0)$ to denote that $\sum_{A \in A_{=}^{(1,0)}} (|A| - 1) = 1$ and the 0 keeps track of this particular instance where $\sum_{A \in A_{=}} (|A| - 1) = 1$).

We next repeat this argument for

$$\sum_{x_1, x_2 \neq \ldots \neq x_r} \Pi_{m=1}^r f(x_m) = \sum_{x_1, x_2, x_3 \neq \ldots \neq x_r} \Pi_{m=1}^r f(x_m) - (r-2) \sum_{x_1, x_2 = x_3 \neq \ldots \neq x_r} \Pi_{m=1}^r f(x_m)$$

and in context to the second summation on the right hand side, define $p_{(1,1)} = -(r-2)$, $A_{=}^{(1,1)} = \{\{2,3\}\}$, the variables that equal one another, and $A_{\neq}^{(1,1)} = \{2, 3, 4, .., r\}$, the variables that are required not to equal some other variable (where we use the superscript $(1, 1)$ to denote that $\sum_{A \in A_{=}^{(1,1)}} (|A| - 1) = 1$ and the 1 keeps track of this new instance where $\sum_{A \in A_{=}} (|A| - 1) = 1$). We stop once we reach

$$\sum_{x_1, x_2, .., x_{2k} \neq \ldots \neq x_r} \Pi_{m=1}^r f(x_m) = \sum_{x_1, \ldots, x_{2k}, x_{2k+1} \neq \ldots \neq x_r} \Pi_{m=1}^r f(x_m) - (r-2k) \sum_{x_1, \ldots, x_{2k} = x_{2k+1} \neq \ldots \neq x_r} \Pi_{m=1}^r f(x_m). \tag{3.44}$$

Finally, based on the the first sum on the right hand side of equation (3.44), we define $p_{(0,0)} = 1$, $A_{=}^{(0,0)} = \emptyset$, as there are no variables that are required to equal one another, $A_{\neq}^{(0,0)} = \{2k+1, ..., r\}$,the variables that are required not to equal some other variable, as $\sum_{A \in A_{=}^{(0,0)}} (|A| - 1) = 0$, and note that only variables with an index of $2k + 1$ or greater are constrained so that they cannot equal one another. Note that in the theorem statement, the number of instances where we have sets of the form $A_{=}^{(1,x)} \otimes A_{\neq}^{(1,x)}$ is $2k$, but since we start with $c = 0$, the maximumum value of c will be $2k - 1$ (and hence define $h(1, k) = 2k - 1$.)

From the above inductive construction, we know that for all $c \in \{0, 1, .., 2k-1\}$,

$$|p_{(1,c)}| \leq r. \tag{3.45}$$

. In addition, since the range of f consists of the non-negative real numbers, we know that for all c,

$$\sum_{A_{=}^{(1,c)} \otimes A_{\neq}^{(1,c)}} \Pi f(x_m) \leq \sum_{A_{=}^{(1,2k-1)} \otimes A_{\neq}^{(1,2k-1)}} \Pi f(x_m) \tag{3.46}$$

as the number of variables that are required not to equal some other variable in $A_{\neq}^{(1,2k-1)}$ is smaller than the analogous list for any other instance of $A_{\neq}^{(1,c)}$. Hence by inequalities (3.45) and (3.46), we find that

$$\frac{|\sum_{c=0}^{2k-1} p_{(1,c)} \sum_{A_{=}^{(1,c)} \otimes A_{\neq}^{(1,c)}} \Pi f(x_m)|}{\sum_{x_1,\ldots,x_{2k},x_{2k+1}\neq\ldots\neq x_r} \Pi f(x_m)} \leq 2kr \frac{\sum_{A_{=}^{(1,2k-1)} \otimes A_{\neq}^{(1,2k-1)}} \Pi f(x_m)}{\sum_{x_1,\ldots,x_{2k},x_{2k+1}\neq\ldots\neq x_r} \Pi f(x_m)} =$$

$$2kr \frac{\sum_{x_1,\ldots,x_{2k-1},x_{2k}=x_{2k+1}\neq\ldots\neq x_r} \Pi f(x_m)}{\sum_{x_1,\ldots,x_{2k},x_{2k+1}\neq\ldots\neq x_r} \Pi f(x_m)} = 2kr \frac{S^{2k-1} \sum_{x_{2k}=x_{2k+1}\neq\ldots\neq x_r} \Pi_{m=2k}^{r} f(x_m)}{S^{2k} \sum_{x_{2k+1}\neq\ldots\neq x_r} \Pi f(x_m)} \leq$$

$$2krf(x_{max}) \frac{\sum_{x_{2k+1}\neq\ldots\neq x_r} \Pi_{m=2k+1}^{r} f(x_m)}{S \sum_{x_{2k+1}\neq\ldots\neq x_r} \Pi f(x_m)} = \frac{2krf(x_{max})}{S} = 2k \frac{O(S^{1-2\tau})}{S} = O(S^{-2\tau})$$

where all products without indices labeled are taken over the set of m values in the index set of the summation that precedes them. Thus, we have attained our desired results for $j = 1$.

To proceed with induction, we now assume that the inductive statement holds for $j = n-1$ and prove that it is true for $j = n$, provided $n \leq k$. The inductive hypothesis when $j = n-1$ yields

$$\sum_{x_1\neq\ldots\neq x_r} \Pi_{m=1}^{r} f(x_m) = \sum_{i=0}^{n-1} \sum_{c=0}^{h(i,k)} p_{(i,c)} \sum_{A_{=}^{(i,c)} \otimes A_{\neq}^{(i,c)}} \Pi_{m=1}^{r} f(x_m).$$

To prove that the inductive statement holds for $j = n$ we manipulate the sets of the form $A_{=}^{(n-1,c)} \otimes A_{\neq}^{(n-1,c)}$ in our inductive hypothesis (where $j = n-1$). From our inductive hypothesis we know that $\sum_{A \in A_{=}^{(n-1,c)}} (|A| - 1) = n - 1$; for some $s \leq 3k$, $A_{\neq}^{(n-1,c)} = \{s, s+$

$1, ..., r\}$; and there exists at most one set in $A_{=}^{(n-1,c)}$, of the form $\{s, s+1, ..., s+t\}$, that has a nontrivial intersection with $A_{\neq}^{(n-1,c)}$.

We proceed as in the base case.

Case 1: Suppose that indeed for some s, t, $\{s, s+1, ..., s+t\} \in A_{=}^{(n-1,c)}$ and $A_{\neq}^{(n-1,c)} = \{s, s+1, ..., r\}$. By assumption we can write $A_{=}^{(n-1,c)} = A_1 \cup \{s, s+1, ..., s+t\}$ where we define A_1 to be $A_{=}^{(n-1,c)} - \{s, s+1, ..., s+t\}$. Since $x_s = x_{s+1} = ... = x_{s+t}$, we have the following equality:

$$p_{(n-1,c)} \sum_{A_{=}^{(n-1,c)} \otimes A_{\neq}^{(n-1,c)}} \Pi f(x_m) = p_{(n-1,c)} \sum_{A_{=}^{(n-1,c)} \otimes \{s+t+1,...,r\}} \Pi f(x_m) - \tag{3.47}$$

$$p_{(n-1,c)}(r-s-t) \sum_{A_1 \cup \{s,...,s+t+1\} \otimes \{s,...,r\}} \Pi f(x_m).$$

The last term on the right hand side of (3.47) is a sum over an index set with n equalities. Analogous to the proof of the base case, we express this index set as $A_{=}^{(n,c_*)} \otimes A_{\neq}^{(n,c_*)}$ for some label c_*. Of course, the product of the two polynomial coefficients of this term is a polynomial as well, which we can take as $p_{(n,c_*)}$. To see how this works, we proceed for each c depending on whether $s+t \leq 2k$ or $s+t > 2k$.

Case 1a: $s+t \leq 2k$

Analogous to the proof in the base case, the expression (3.47) can be decomposed as

$$\sum_{\substack{A_{=}^{(n-1,c)} \otimes \\ \{s+t+1,...,r\}}} \Pi f(x_m) = \sum_{\substack{A_{=}^{(n-1,c)} \otimes \\ \{s+t+2,...,r\}}} \Pi f(x_m) - (r-s-t-1) \sum_{\substack{A_{=}^{(n-1,c)} \cup \{s+t+1,s+t+2\} \otimes \\ \{s+t+1,...,r\}}} \Pi f(x_m).$$

We use the last term to construct $A_{=}^{(n,c_*)} \otimes A_{\neq}^{(n,c_*)}$ for some c_*. (Note that $\sum_{A \in A_{=}^{(n,c_*)}} (|A|-1) = n$). We continue this process until we reach

$$\sum_{A_{=}^{(n-1,c)} \otimes \{2k,...,r\}} \Pi f(x_m) = \sum_{A_{=}^{(n-1,c)} \otimes \{2k+1,...,r\}} \Pi f(x_m) - \tag{3.48}$$

$$(r-2k) \sum_{A_{=}^{(n-1,c)} \cup \{2k,2k+1\} \otimes \{2k,...,r\}} \Pi f(x_m).$$

We note that with $j = n$, the $i = n-1$ terms in (3.43) may be different from the $i = n-1$ terms with $j = n-1$. Indeed, based on (3.48), we now redefine (for $j = n$) $A_{=}^{(n-1,c)} \otimes A_{\neq}^{(n-1,c)} = A_{=}^{(n-1,c)} \otimes \{2k+1, ..., r\}$.

Case 1b: $s + t > 2k$

Now since by the assumption of the inductive hypothesis, $\sum_{A \in A_{=}^{(n-1,c)}} (|A| - 1) = n - 1 \leq k - 1 \implies \sum_{A \in A_{=}^{(n-1,c)}} |A| \leq 2k - 2$, there are at least $s + t - 2k + 2$ values in $\{1, ..., s+t\}$ that are not in A for all $A \in A_{=}^{(n-1,c)}$. We then create a relabeling where all values greater than $s + t$ are left alone (they map to themselves) and that the values that are not in A for all $A \in A_{=}^{(n-1,c)}$ include $\{2k+1, ..., s+t\}$. (Hence the largest element in all of the sets in $A_{=}^{(n-1,c)}$ is bounded by $2k$.)

Expression (3.47) can be decomposed as

$$\sum_{A_{=}^{(n-1,c)} \otimes \{s+t+1,...,r\}} \Pi f(x_m) = \sum_{A_{=}^{(n-1,c)} \otimes \{s+t,...,r\}} \Pi f(x_m) + (r-s-t) \sum_{A_{=}^{(n-1,c)} \cup \{s+t,s+t+1\} \otimes \{s+t,...,r\}} \Pi f(x_m)$$

As in Case 1a, we use the last term to construct $A_{=}^{(n,c_*)} \otimes A_{\neq}^{(n,c_*)}$ for some c_*. We repeat this equality until we attain

$$\sum_{A_{=}^{(n-1,c)} \otimes \{2k+2,...,r\}} \Pi f(x_m) = \sum_{A_{=}^{(n-1,c)} \otimes \{2k+1,...,r\}} \Pi f(x_m) + (r-2k-1) \sum_{A_{=}^{(n-1,c)} \cup \{2k,2k+1\} \otimes \{2k,...,r\}} \Pi f(x_m)$$

Case 2: Suppose that $A_{\neq}^{(n-1,c)} = \{s, s+1, ..., r\}$ and for all t, $\{s, s+1, ..., s+t\} \notin A_{=}^{(n-1,c)}$. We again consider two subcases.

Case 2a: $s \leq 2k$.

To achieve the desired form, we rewrite

$$\sum_{A_{=}^{(n-1,c)} \otimes \{s,...,r\}} \Pi f(x_m) = \sum_{A_{=}^{(n-1,c)} \otimes \{s+1,...,r\}} \Pi f(x_m) - (r-s-1) \sum_{A_{=}^{(n-1,c)} \cup \{s,s+1\} \otimes \{s,...,r\}} \Pi f(x_m).$$

We use the last term to construct $A_{=}^{(n,c_*)} \otimes A_{\neq}^{(n,c_*)}$ for some c_*, proceeding inductively until we reach

$$\sum_{A_{=}^{(n-1,c)} \otimes \{2k,\ldots,r\}} \Pi f(x_m) = \sum_{A_{=}^{(n-1,c)} \otimes \{2k+1,\ldots,r\}} \Pi f(x_m) - (r-2k-1) \sum_{A_{=}^{(n-1,c)} \cup \{2k,2k+1\} \otimes \{2k,\ldots,r\}} \Pi f(x_m).$$

Case 2b: $s > 2k$.

If $A_{=}^{(n-1,c)}$ contains sets with elements that are greater than $2k$, we can again perform a relabeling scheme as in Case 1b such that all of the elements in $A_{=}^{(n-1,c)}$ are guaranteed to be bounded above by $2k$:

$$\sum_{A_{=}^{(n-1,c)} \otimes \{s,\ldots,r\}} \Pi f(x_m) = \sum_{A_{=}^{(n-1,c)} \otimes \{s-1,\ldots,r\}} \Pi f(x_m) + (r-s-1) \sum_{A_{=}^{(n-1,c)} \cup \{s-1,s\} \otimes \{s-1,\ldots,r\}} \Pi f(x_m).$$

We use the last term to construct $A_{=}^{(n,c_*)} \otimes A_{\neq}^{(n,c_*)}$ for some c_* and we repeat this equality until we attain

$$\sum_{A_{=}^{(n-1,c)} \otimes \{2k+2,\ldots,r\}} \Pi f(x_m) = \sum_{A_{=}^{(n-1,c)} \otimes \{2k+1,\ldots,r\}} \Pi f(x_m) + (r-2k-1) \sum_{A_{=}^{(n-1,c)} \cup \{2k,2k+1\} \otimes \{2k,\ldots,r\}} \Pi f(x_m).$$

To complete the proof, we need to verify the following properties from our theorem statement:

- $A_{=}^{(i,c)}$ is a collection of disjoint subsets of $\{1,\ldots,2k\}$. This fact follows from the conclusion of each of the cases in the above inductive argument.
- $\sum_{A \in A_{=}^{(i,c)}} (|A|-1) = i$. For each $A_{=}^{(i,c_*)}$ constructed from an $A_{=}^{(i-1,c)}$, it is easy to show that $\sum_{A \in A_{=}^{(i-1,c)}} (|A|-1) = i-1$, and then the inductive step adds two more terms to the sum to yield $\sum_{A \in A_{=}^{(i,c_*)}} (|A|-1) = i$.
- For $i < j$, $A_{\neq}^{(i,c)} = \{2k+1,...,r\}$. This fact is easily verified (and follows trivially) from the construction through the inductive hypothesis.
- $A_{\neq}^{(j,c)} = \{s, s+1, ..., r\}$ for some $s \leq 3k$. This fact is also easily verified (and follows trivially) from the proof by induction above.
- If $A \in A_{=}^{(j,c)}$ has a nontrivial intersection with $A_{\neq}^{(j,c)}$ then $A = \{s, s+1, ..., s+t\}$; in addition, there can only be one such $A \in A_{=}^{(j,c)}$ that has a nontrivial intersection with $A_{\neq}^{(j,c)}$. This fact is also easily verified (and follows trivially) from the proof by induction above.

- $$\frac{p_{(i,c)} \sum_{A_{=}^{(i,c)} \otimes A_{\neq}^{(i,c)}} \Pi_{m=1}^{r} f(x_m)}{\sum_{x_1,\ldots,x_{2k},x_{2k+1} \neq \ldots \neq x_r} \Pi_{m=1}^{r} f(x_m)} = O(S^{-2i\tau}). \tag{3.49}$$

The proof for this result is tricky but also follows from the inductive argument. In the inductive hypothesis, assume for all sets $A_{=}^{(i-1,c)} \otimes A_{\neq}^{(i-1,c)}$ and corresponding polynomials $p_{(i-1,c)}$ that equation (3.49) holds. Consider what happens when we construct sets of the form $A_{=}^{(i,c_*)} \otimes A_{\neq}^{(i,c_*)}$ and the corresponding polynomials $p_{(i,c_*)}$ from $A_{=}^{(i-1,c)} \otimes A_{\neq}^{(i-1,c)}$. We know from the above inductive proof that $p_{(i,c_*)} \leq r p_{(i-1,c)}$. Furthermore, there is an additional equal sign under the summation with the index set $A_{=}^{(i,c_*)} \otimes A_{\neq}^{(i,c_*)}$. Suppose that x_m is a variable that does not appear in $A_{=}^{(i-1,c)} \otimes A_{\neq}^{(i-1,c)}$ but does appear in $A_{=}^{(i,c_*)} \otimes A_{\neq}^{(i,c_*)}$. We bound the contribution of $f(x_m)$ in $A_{=}^{(i,c_*)} \otimes A_{\neq}^{(i,c_*)}$ by $d_{max} = O(S^{\frac{1}{2}-\tau})$. That is,

$$\frac{p_{(i,c_*)} \sum_{A_{=}^{(i,c_*)} \otimes A_{\neq}^{(i,c_*)}} \Pi f(x_k)}{p_{(i-1,c)} \sum_{A_{=}^{(i-1,c)} \otimes A_{\neq}^{(i-1,c)}} \Pi f(x_k)} \leq \frac{d_{max} r \sum_{A_{=}^{(i,c_*)} \otimes A_{\neq}^{(i,c_*)}} \Pi_{k \neq m} f(x_k)}{\sum_{A_{=}^{(i-1,c)} \otimes A_{\neq}^{(i-1,c)}} \Pi f(x_k)}. \tag{3.50}$$

Now that $f(x_m)$ has been 'taken out', we can construct a crude lower bound on the contribution of $f(x_m)$ in $A_{=}^{(i-1,c)} \otimes A_{\neq}^{(i-1,c)}$, given the hypothesis that $\sum_{x_m} f(x_m) = O(S)$. So we have from the inequality (3.50) that

$$\frac{p_{(i,c_*)} \sum_{A_{=}^{(i,c_*)} \otimes A_{\neq}^{(i,c_*)}} \Pi f(x_k)}{p_{(i-1,c)} \sum_{A_{=}^{(i-1,c)} \otimes A_{\neq}^{(i-1,c)}} \Pi f(x_k)} \leq \frac{O(S^{1-2\tau}) \sum_{A_{=}^{(i,c_*)} \otimes A_{\neq}^{(i,c_*)}} \Pi_{k \neq m} f(x_k)}{O(S) \sum_{A_{=}^{(i-1,c)} \otimes A_{\neq}^{(i-1,c)}} \Pi_{k \neq m} f(x_k)}. \tag{3.51}$$

But now that x_m has been effectively removed, the summations over $A_{=}^{(i-1,c)} \otimes A_{\neq}^{(i-1,c)}$ and $A_{=}^{(i,c_*)} \otimes A_{\neq}^{(i,c_*)}$ are identical and we conclude from the right hand side of (3.51) that

$$\frac{p_{(i,c_*)} \sum_{A_{=}^{(i,c_*)} \otimes A_{\neq}^{(i,c_*)}} \Pi_{k \neq m} f(x_k)}{p_{(i-1,c)} \sum_{A_{=}^{(i-1,c)} \otimes A_{\neq}^{(i-1,c)}} \Pi_{k \neq m} f(x_k)} \leq O(S^{-2\tau}) \tag{3.52}$$

Invoking the inductive hypothesis that

$$\frac{p_{(i-1,c)} \sum_{A_{=}^{(i-1,c)} \otimes A_{\neq}^{(i-1,c)}} \Pi_{m=1}^{r} f(x_m)}{\sum_{x_1,\ldots,x_{2k},x_{2k+1} \neq \ldots \neq x_r} \Pi_{m=1}^{r} f(x_m)} = O(S^{-2(i-1)\tau}), \tag{3.53}$$

and combining this with inequalities (3.51),(3.52) completes the proof.

□

We now prove the desired result that validates the assumptions in Theorems 9 and 10 that enable us to extend our enumeration estimate to arbitrary order. But before doing so, we would like to remind the reader that throughout this work we have been expressing our approximation f of ϕ such that $f = \phi * (1 + O(S^{-2k\tau}))$. This statement implies that $f - \phi = \phi O(S^{-2k\tau})$ But since we know by Corollary 8 that $\phi(x_i, x_j, \mathbf{d}, \mathbf{a}) = \frac{a_i}{a_j}(1 + O(S^{-2\tau}))$, it follows that $f - \phi = \frac{a_i}{a_j} O(S^{-2k\tau})$. In the theorem that follows we often interchange $f = \phi(1 + O(S^{-2k\tau})$. $f - \phi = \phi O(S^{-2k\tau})$ and $f - \phi = \frac{a_i}{a_j} O(S^{-2k\tau})$.

Theorem 18. *Given an approximation f of ϕ (the true value for the ratio of the number of graphs that realize two slightly different degree sequences) such that $|f(x_i, x_j, \mathbf{d}) - \phi(x_i, x_j, \mathbf{d})| = \frac{a_i}{a_j} O(S^{-2k\tau})$ for some $k < \frac{1}{4\tau} + 1$,*

$$f(x_i, x_j, \mathbf{d}, \mathbf{a}) = \frac{h(x_i, \mathbf{d}, \mathbf{a})}{h(x_j, \mathbf{d}, \mathbf{a})} \tag{3.54}$$

where

$$h(x_i, \mathbf{d}, \mathbf{a}) = a_i(1 + \sum_{v=1}^{r} \frac{\gamma_v a_i^{k_v} \prod_{q=1}^{s} \beta_q^{m(v,q)} \alpha_q^{n(v,q)}}{\alpha_1^z}) \tag{3.55}$$

for some constants γ_v,where r, s, z are finite and each term in the summations in the numerator and denominator in (3.54) is $O(S^{-2\tau})$.

Furthermore if $\|\mathbf{d}_1 - \mathbf{d}_2\|_\infty \leq 1$ and $\|\mathbf{d}_1 - \mathbf{d}_2\|_1 = O(S^{\frac{1}{2}-\tau})$, assuming the degrees of node x_i are identical in $\mathbf{d}_1$ and $\mathbf{d}_2$, then

$$|h(x_i, \mathbf{d}_1) - h(x_i, \mathbf{d}_2)| = a_i(O(S^{-\frac{1}{2}-3\tau})).$$

Proof. We prove the result by induction.

The base case holds trivially when $k = 1$ as by Corollary 9, $f(x_i, x_j, \mathbf{d}) = a_i/a_j$ and it also follows trivially that differences in the degree sequence aside from the degrees of nodes x_i or x_j do not affect $f(x_i, x_j, \mathbf{d})$. Similarly, equation (3.54) holds with $r = 1$ and $\gamma_1 = 0$ in the expression (3.55) for both the numerator and denominator of (3.54). So we proceed to the inductive step.

Given our inductive hypothesis where we have an approximation f where the dependence on the degree sequence in f is sufficiently weak $O(S^{-\frac{1}{2}-\tau})$ and can be dropped as

$|f(x_i,x_j,\mathbf{d})-\phi(x_i,x_j,\mathbf{d})| = \frac{a_i}{a_j}(O(S^{-2(k-1)\tau}))$ since $2(k-1)\tau < \frac{1}{2}+2\tau$, we want to show that the sharper approximation produced by Theorem 9 also has the same property.

By applying Theorem 9, we get a stronger approximation (which we also denote by f) which is of the form $\frac{a_i}{a_j}\exp(\log(1+\frac{\|G_{X_{1i}}\|}{\|G_{X_{0i}}\|}))\exp(-\log(1+\frac{\|G_{X_{1j}}\|}{\|G_{X_{0j}}\|}))$. From equations (3.14)-(3.16) in Theorem 9, we have

$$\frac{\|G_{X_{1i}}\|}{\|G_{X_{0i}}\|} = \frac{(a_i-1)\sum_{x_1\neq\ldots\neq x_{a_i-1}=x_{a_i}}\Pi f(x_i,m_i,\mathbf{d},\mathbf{b})}{\sum_{x_1\neq\ldots\neq x_{a_i}}\Pi f(x_i,m_i,\mathbf{d},\mathbf{b})}$$

where the choice of m_i is arbitrary. By the inductive hypothesis (3.54), we can drop the dependence on m_i by multiplying the numerator and denominator by $\Pi_{i=1}^{a_i} b_{m_i}(1+\sum_{v=1}^{r}\frac{\gamma_v b_{m_i}^{k_v}\Pi_{q=1}^{s}\beta_q^{m(v,q)}\alpha_q^{n(v,q)}}{\alpha_1^z})$; that is, we have

$$\frac{\|G_{X_{1i}}\|}{\|G_{X_{0i}}\|} = \frac{(a_i-1)\sum_{x_1\neq\ldots\neq x_{a_i-1}=x_{a_i}}\Pi f(x_i,\mathbf{d},\mathbf{b})}{\sum_{x_1\neq\ldots\neq x_{a_i}}\Pi f(x_i,\mathbf{d},\mathbf{b})}$$

where

$$f(x_i,\mathbf{d},\mathbf{b}) = b_i(1+\sum_{v=1}^{r}\frac{\gamma_v b_i^{k_v}\Pi_{q=1}^{s}\beta_q^{m(v,q)}\alpha_q^{n(v,q)}}{\alpha_1^z}). \tag{3.56}$$

For simplicity, we will now drop the explicit dependence on $\mathbf{d}$ and $\mathbf{b}$ from f. Note that by applying Theorems 16 and 17 to the denominator and numerator respectively we have that

$$\frac{\|G_{X_{1i}}\|}{\|G_{X_{0i}}\|} = \frac{(a_i-1)\sum_{j=0}^{k}\sum_{c=0}^{h(j,k)}p_{(j,c,1)}(a_i)\sum_{A_=^{(j,c,1)}\otimes A_{\neq}^{(j,c,1)}}\Pi f(x_i)}{\sum_{j=0}^{k}\sum_{c=0}^{h(j,k)}p_{(j,c,0)}(a_i)\sum_{A_=^{(j,c,0)}\otimes A_{\neq}^{(j,c,0)}}\Pi f(x_i)}$$

where the p's are polynomials in a_i, summations denoted with $A_=^{(j,c,x)}$ are over a finite number $2k$ of variables. But by Theorems 16 and 17, we can drop the terms in the numerator and denominator involving $p_{(k,c,x)}\sum_{A_=^{(k,c,x)}\otimes A_{\neq}^{(k,c,x)}}\Pi f(x_i)$ as they only contribute a maximum of $O(S^{-2k\tau})$, hence

$$\frac{(a_i-1)\sum_{j=0}^{k-1}\sum_{c=0}^{h(j,k)}p_{(j,c,1)}(a_i)\sum_{A_=^{(j,c,1)}\otimes A_{\neq}^{(j,c,1)}}\Pi f(x_i)}{\sum_{j=0}^{k-1}\sum_{c=0}^{h(j,k)}p_{(j,c,0)}(a_i)\sum_{A_=^{(j,c,0)}\otimes A_{\neq}^{(j,c,0)}}\Pi f(x_i)}(1+O(S^{-2k\tau})).$$

(As an abuse of notation, when we write $\sum_{A_=}$, we are summing over all $\mathbf{x}\in\mathbb{N}^{2k}$, but when we consider $A_=$ without the sigma, we are specifying which dummy variables must equal one another.) In addition, by Theorems 16 and 17, for all $j \le k-1$, $\cup_{A\in A_=^{(j,c,x)}}A\cap A_{\neq}^{(j,c,x)} = \emptyset$

and $A_{\neq}^{(j,c,x)} = \{2k+1, ..., a_i\}$. Hence we can factor out a $\sum_{x_{2k+1}\neq \ldots \neq x_{a_i}} \Pi_{i=2k+1}^{a_i} f(x_i)$ from both the numerator and denominator. This yields

$$\frac{\|G_{X_{1i}}\|}{\|G_{X_{0i}}\|} = \frac{(a_i - 1)\sum_{j=0}^{k-1}\sum_{c=0}^{h(j,k)} p_{(j,c,1)}(a_i)\sum_{A_{=}^{(j,c,1)}} \Pi f(x_i)}{\sum_{j=0}^{k-1}\sum_{c=0}^{h(j,k)} p_{(j,c,0)}(a_i)\sum_{A_{=}^{(j,c,0)}} \Pi f(x_i)} + O(S^{-2k\tau})).$$

Recall that $f(x_i, \mathbf{d}, \mathbf{b}) = b_i(1 + \sum_{v=1}^{r} \frac{\gamma_v b_i^{k_v} \Pi_{q=1}^{s} \beta_q^{m(v,q)} \alpha_q^{n(v,q)}}{\alpha_1^z})$ where each of the finitely many terms in the summation is $O(S^{-2\tau})$. Furthermore, note that only $A_{=}^{(0,0,0)} = \emptyset$, so let us multiply the numerator and denominator by $\frac{1}{\alpha^{2k}}$:

$$\frac{\|G_{X_{1i}}\|}{\|G_{X_{0i}}\|} = \frac{(a_i - 1)\sum_{j=0}^{k-1}\sum_{c=0}^{h(j,k)} p_{(j,c,1)}(a_i)\sum_{A_{=}^{(j,c,1)}} \frac{\Pi f(x_i)}{\alpha_1^{2k}}}{\sum_{j=0}^{k-1}\sum_{c=0}^{h(j,k)} p_{(j,c,0)}(a_i)\sum_{A_{=}^{(j,c,0)}} \frac{\Pi f(x_i)}{\alpha_1^{2k}}} + O(S^{-2k\tau}). \tag{3.57}$$

Note that every term in the numerator $(a_i - 1)p_{(j,c,1)}(a_i)\sum_{A_{=}^{(j,c,1)}} \frac{\Pi f(x_i)}{\alpha_1^{2k}} = O(S^{-2\tau})$ and except for $j = 0$ and $c = 0$, $p_{(j,c,0)}(a_i)\sum_{A_{=}^{j,c,0)}} \frac{\Pi f(x_i)}{\alpha_1^{2k}} = O(S^{-2\tau})$. And finally, since $A_{=}^{(0,0,0)} = \emptyset$ and $f(x_i, \mathbf{d}, \mathbf{b}) = b_i(1 + \sum_{v=1}^{r} \frac{\gamma_v b_i^{k_v} \Pi_{q=1}^{s} \beta_q^{m(v,q)} \alpha_q^{n(v,q)}}{\alpha_1^z})$ where each term in the summation is $O(S^{-2\tau})$and $\sum f(x_i) = S + O(S^{1-2\tau})$, we note that $\sum_{A_{=}^{(0,0,0)}} \frac{\Pi f(x_i)}{\alpha_1^{2k}} = 1 + O(S^{-2\tau})$ as $\alpha_1 = S$. We now show that the dependence in equation (3.57) on the degree sequence is 'small'.

As noted before, $(a_i - 1)^{\delta_{x,1}} p_{(j,c,x)}(a_i)\sum_{A_{=}^{(j,c,x)}} \frac{\Pi f(x_i)}{\alpha_1^{2k}} = \delta_{x,0} + O(S^{-2\tau})$ where $\delta_{x,y} = 1$ if $x = y$ and 0 otherwise and the $O(S^{-2\tau})$ term is some finite sum of terms of the form $\gamma a_i^j \frac{\Pi_{k=1}^m \alpha_k^{g(k)} \beta_k^{h(k)}}{\alpha_1^{2k}}$ each of which are $O(S^{-2\tau})$.

The constraint that $\gamma a_i^j \frac{\Pi_{k=1}^m \alpha_k^{g(k)} \beta_k^{h(k)}}{\alpha_1^{2k}} = O(S^{-2\tau})$ holds for all degree sequences such that $d_{max} = O(S^{\frac{1}{2}-\tau})$. Now we consider the perturbation analysis, such that $\|\mathbf{d}_1 - \mathbf{d}_0\|_1 = O(S^{\frac{1}{2}-\tau})$ and $\|\mathbf{d}_1 - \mathbf{d}_0\|_\infty \leq 1$, then $|\beta_j(\mathbf{d}_1) - \beta_j(\mathbf{d}_0)| = O(d_{max}^j) = O(S^{\frac{j}{2}-j\tau})$. But also note that $\max(\alpha_j, \beta_j) = O(S^{1+\frac{j-1}{2}-(j-1)\tau})$ as $\beta_j = \sum_i b_i^j \leq d_{max}^{j-1}\sum_i b_i = O(S^{1+\frac{j-1}{2}-(j-1)\tau})$. Now if we measure the impact of considering different degree sequences $\mathbf{d}_0, \mathbf{d}_1$ on $\gamma a_i^j \frac{\Pi_{k=1}^m \alpha_k^{g(k)} \beta_k^{h(k)}}{\alpha_1^{2k}}$, this results in a $O(S^{-\frac{1}{2}-\tau})$ times smaller than $\gamma a_i^j \frac{\Pi_{k=1}^m \alpha_k^{g(k)} \beta_k^{h(k)}}{\alpha_1^{2k}}$. But since $\gamma a_i^j \frac{\Pi_{k=1}^m \alpha_k^{g(k)} \beta_k^{h(k)}}{\alpha_1^{2k}} = O(S^{-2\tau})$, the contribution in equation (3.57) from considering different degree sequences is $O(S^{-\frac{1}{2}-\tau-2\tau})$.

Now to verify that our new higher order approximation $f(x_i, x_j, \mathbf{d}, \mathbf{a}) = \frac{a_i}{a_j}\left(\frac{1+\sum_{v=1}^{r} \frac{\gamma_v a_i^{k_v} \Pi_{q=1}^{s} \beta_q^{m(v,q)} \alpha_q^{n(v,q)}}{\alpha_1^z}}{1+\sum_{v=1}^{r} \frac{\gamma_v a_j^{k_v} \Pi_{q=1}^{s} \beta_q^{m(v,q)} \alpha_q^{n(v,q)}}{\alpha_1^z}}\right)$, we merely perform a Taylor expansion in the denominator of

$$\frac{(a_i-1)\sum_{j=0}^{k-1}\sum_{c=0}^{h(j,k)} p_{(j,c,1)}(a_i)\sum_{A_{=}^{(j,c,1)}}\frac{\Pi f(x_i)}{\alpha_1^{2k}}}{\sum_{j=0}^{k-1}\sum_{c=0}^{h(j,k)} p_{(j,c,0)}(a_i)\sum_{A_{=}^{(j,c,0)}}\frac{\Pi f(x_i)}{\alpha_1^{2k}}}+O(S^{-2k\tau})$$

as the denominator can be rewritten in the form $1+O(S^{-2\tau})$ where the $O(S^{-2\tau})$ term is some finite sum of terms of the form $\gamma a_i^j \frac{\Pi_{k=1}^m \alpha_k^{g(k)}\beta_k^{h(k)}}{\alpha_1^{2k}}$ each of which are $O(S^{-2\tau})$ and each term in the numerator is $O(S^{-2\tau})$. Hence we get that

$$1+\frac{(a_i-1)\sum_{j=0}^{k-1}\sum_{c=0}^{h(j,k)} p_{(j,c,1)}(a_i)\sum_{A_{=}^{(j,c,1)}}\frac{\Pi f(x_i)}{\alpha_1^{2k}}}{\sum_{j=0}^{k-1}\sum_{c=0}^{h(j,k)} p_{(j,c,0)}(a_i)\sum_{A_{=}^{(j,c,0)}}\frac{\Pi f(x_i)}{\alpha_1^{2k}}}+O(S^{-2k\tau})=(1+\sum_{v=1}^{r_*}\frac{\gamma_v a_i^{k_v}\Pi_{q=1}^{s}\beta_q^{m(v,q)}}{\alpha_1^{2k}})$$

for some finite r_*.

Repeating the same argument for evaluating the $\exp(-\log(1+\frac{\|G_{X1j}\|}{\|G_{X0j}\|}))$ term yields the desired result, that $f(x_i,x_j,\mathbf{d},\mathbf{a})=\frac{a_i}{a_j}\left(\frac{1+\sum_{v=1}^{r}\frac{\gamma_v a_i^{k_v}\Pi_{q=1}^{s}\beta_q^{m(v,q)}\alpha_q^{n(v,q)}}{\alpha_1^{z}}}{1+\sum_{v=1}^{r}\frac{\gamma_v a_j^{k_v}\Pi_{q=1}^{s}\beta_q^{m(v,q)}\alpha_q^{n(v,q)}}{\alpha_1^{z}}}\right)$.

□

3.6 ENUMERATING GRAPHS WITH GREATER DEPENDENCE ON THE DEGREE SEQUENCE

In this section, we provide a proof of Theorem 10. From Theorem 9, we have an approximation f such that $\frac{\|G_{\mathbf{d}_{-\mathbf{i}}}\|}{\|G_{\mathbf{d}_{-\mathbf{j}}}\|}=f(\mathbf{e}_i,\mathbf{e}_j,\mathbf{d},\sigma)(1+O(S^{-\frac{1}{2}-w\tau}))$ where $\sigma=\mathbf{a}$ or $\mathbf{b}$, where we have assumed (and justified) up until this point that the dependence on the degree sequence $\mathbf{d}$ in f is weak. Eventually, we will reach a point where such an assumption is no longer reasonable. In this case we decompose our approximation f into one part that effectively ignores the dependence of $\mathbf{d}$ and a part of f that takes into account $\mathbf{d}$. That is, we can write

$$f(\mathbf{e}_i,\mathbf{e}_j,\mathbf{d}_1,\sigma)=f(\mathbf{e}_i,\mathbf{e}_j,\mathbf{d}_0,\sigma)(1+O(S^{-\frac{1}{2}-\tau})), \tag{3.58}$$

where we have proven that for suitably chosen $\mathbf{d}_0$ the correction will yield a term of size at most $O(S^{-\frac{1}{2}-\tau})$. For simplicity, we will consider $\sigma = \mathbf{a}$. Denote the $O(S^{-\frac{1}{2}-\tau})$ term in equation (3.58) by z_*, such that equation (3.58) becomes

$$f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}_1, \mathbf{a}) = f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}_0, \mathbf{a}) + z_* f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}_0, \mathbf{a}). \tag{3.59}$$

In the Theorem below, we restate Theorem 10; note that we implicitly define $z = z_* f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}_0, \mathbf{a})$.

Appendix Theorem 1 (Theorem 10)**.** *Consider an approximation*

$$\frac{\|G_{\mathbf{d_{-i}}}\|}{\|G_{\mathbf{d_{-j}}}\|} = f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}, \sigma)(1 + O(S^{-\frac{1}{2}-w\tau}))$$

for some $w > 0$. Furthermore suppose that for $m = O(S^{\frac{1}{2}-\tau})$,

$$f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}_0, \sigma) = f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}_1, \sigma) + z(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}_0 - \mathbf{d}_1, \mathbf{d}_0, \sigma)$$

where $\|\mathbf{d}_1 - \mathbf{d}_0\|_1 \leq m$ and $z(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}_0 - \mathbf{d}_1, \mathbf{d}_0, \sigma) \leq O(S^{-\frac{1}{2}-\tau}) f(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}_0, \sigma)$. If $\sigma = \mathbf{a}$, then we can construct a sharper approximation

$$\frac{\|G_{\mathbf{d_{-i}}}\|}{\|G_{\mathbf{d_{-j}}}\|} = g(\mathbf{e}_i, \mathbf{e_j}, \mathbf{d}, \mathbf{a})(1 + O(S^{-\frac{1}{2}-(w+2)\tau}))$$

where

$$g(\mathbf{e}_i, \mathbf{e}_j, \mathbf{d}, \mathbf{a}) =$$

$$\frac{a_i}{a_j} \exp(\log(1 + \frac{(a_i - 1)\sum_{x_1 \neq \ldots \neq x_{a_{i-1}} = x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, (\mathbf{a} - a_i\mathbf{e}_i, (\mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} + \sum_{j=1}^{a_i - k} \mathbf{e}_{u_j}), \mathbf{b})}{\sum_{x_1 \neq \ldots \neq x_{a_{i-1}} \neq x_{a_i}} \Pi_{k=1}^{a_i} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, (\mathbf{a} - a_i\mathbf{e}_i, \mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} + \sum_{j=1}^{a_i - k} \mathbf{e}_{u_j}), \mathbf{b})}) -$$

$$\log(1 + \frac{(a_j - 1)\sum_{x_1 \neq \ldots \neq x_{a_{j-1}} = x_{a_j}} \Pi_{k=1}^{a_j} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, (\mathbf{a} - a_j e_j, \mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} + \sum_{j=1}^{a_i - k} \mathbf{e}_{u_j}), \mathbf{b})}{\sum_{x_1 \neq \ldots \neq x_{a_{j-1}} \neq x_{a_j}} \Pi_{k=1}^{a_j} f(\mathbf{e}_{x_k}, \mathbf{e}_{u_k}, (\mathbf{a} - a_j e_j, \mathbf{b} - \sum_{j=1}^{k-1} \mathbf{e}_{x_j} + \sum_{j=1}^{a_i - k} \mathbf{e}_{u_j}), \mathbf{b})}))$$

for an arbitrary choice of u_k. A similar result holds, with g depending on $\mathbf{b}$, if $\sigma = \mathbf{b}$.

Proof. Recall the notation from Theorem 9. As before, we define Δ_i to be the difference between $\frac{\|G_{X_{1_i}}\|}{a_i\|G_{X_{0_i}}\|}$ evaluated using the exact ratio ϕ and the same quantity evaluated using the approximation f. In essence, we will show that using our crude approximation f will yield an approximation of $\frac{\|G_{X_{1_i}}\|}{a_i\|G_{X_{0_i}}\|}$ by a factor of $(1+O(S^{-\frac{1}{2}-(w+2)\tau}))$, after which application of Corollary 10 yields the result. Again we consider equation (3.17), that is

$$\Delta_i = \frac{(a_i-1)\sum_{x_1\neq\ldots\neq x_{a_{i-1}}=x_{a_i}}\Pi_{k=1}^{a_i}\phi(\mathbf{e}_{x_k},\mathbf{e}_{u_{a_i-k+1}},(\mathbf{a}-a_i\mathbf{e}_i-\mathbf{e}_j,\mathbf{b}-\sum_{j=1}^{k-1}\mathbf{e}_{x_j}-\sum_{j=1}^{a_i-k}\mathbf{e}_{u_j}),\mathbf{b})}{\sum_{x_1\neq\ldots\neq x_{a_{i-1}}\neq x_{a_i}}\Pi_{k=1}^{a_i}\phi(\mathbf{e}_{x_k},\mathbf{e}_{u_{a_i-k+1}},(\mathbf{a}-a_i\mathbf{e}_i-\mathbf{e}_j,\mathbf{b}-\sum_{j=1}^{k-1}\mathbf{e}_{x_j}-\sum_{j=1}^{a_i-k}\mathbf{e}_{u_j}),\mathbf{b})} -$$

$$\frac{(a_i-1)\sum_{x_1\neq\ldots\neq x_{a_{i-1}}=x_{a_i}}\Pi_{k=1}^{a_i}f(\mathbf{e}_{x_k},\mathbf{e}_{u_{a_i-k+1}},(\mathbf{a}-a_i\mathbf{e}_i-\mathbf{e}_j,(\mathbf{b}-\sum_{j=1}^{k-1}\mathbf{e}_{x_j}-\sum_{j=1}^{a_i-k}\mathbf{e}_{u_j}),\mathbf{b})}{\sum_{x_1\neq\ldots\neq x_{a_{i-1}}\neq x_{a_i}}\Pi_{k=1}^{a_i}f(\mathbf{e}_{x_k},\mathbf{e}_{u_{a_i-k+1}},(\mathbf{a}-a_i\mathbf{e}_i-\mathbf{e}_j,(\mathbf{b}-\sum_{j=1}^{k-1}\mathbf{e}_{x_j}-\sum_{j=1}^{a_i-k}\mathbf{e}_{u_j}),\mathbf{b})}. \tag{3.60}$$

As in the proof of Theorem 9, we abuse notation by using $f_k(x_k)$, $\phi_k(x_k)$ in place of the full expressions for f and ϕ, even though f_k and $\phi_k(x_k)$ do also depend on $x_1, ..., x_{k-1}$. This reduces to a more tractable (but slightly misleading) notation:

$$\Delta_i = \frac{(a_i-1)\sum_{x_1\neq\ldots\neq x_{a_{i-1}}=x_{a_i}}\Pi_{k=1}^{a_i}\phi_k(x_k)}{\sum_{x_1\neq\ldots\neq x_{a_{i-1}}\neq x_{a_i}}\Pi_{k=1}^{a_i}\phi_k(x_k)} - \frac{(a_i-1)\sum_{x_1\neq\ldots\neq x_{a_{i-1}}=x_{a_i}}\Pi_{k=1}^{a_i}f_k(x_k)}{\sum_{x_1\neq\ldots\neq x_{a_{i-1}}\neq x_{a_i}}\Pi_{k=1}^{a_i}f_k(x_k)}.$$

Let D_0 denote the set of sets of a_i indices in $\{1,\ldots,N\}$ such that the variables associated with these indices are distinct and let D_1 denote the set of sets of a_i indices in $\{1,\ldots,N\}$ such that the variables associated with the first a_i-2 are distinct and those with the final two are equal. Writing Δ_i as a single fraction, we obtain

$$\begin{aligned}\Delta_i &= \frac{(a_i-1)[\sum_{D_1}\Pi_{k=1}^{a_i}\phi_k(x_k)\sum_{D_0}\Pi_{k=1}^{a_i}f_k(x_k)-\sum_{D_1}\Pi_{k=1}^{a_i}f_k(x_k)\sum_{D_0}\Pi_{k=1}^{a_i}\phi_k(x_k)]}{\sum_{D_0}\Pi_{k=1}^{a_i}\phi_k(x_k)\sum_{D_0}\Pi_{k=1}^{a_i}f_k(x_k)} \\ &= \frac{(a_i-1)[\sum_{D_1,D_0}(\Pi_{x_k\in D_1}\phi_k(x_k)\Pi_{x_k\in D_0}f_k(x_k)-\Pi_{x_k\in D_1}f_k(x_k)\Pi_{x_k\in D_0}\phi_k(x_k))]}{\sum_{D_0}\Pi_{k=1}^{a_i}\phi_k(x_k)\sum_{D_0}\Pi_{k=1}^{a_i}f_k(x_k)}.\end{aligned}$$

check: We now apply Theorem 18 repeatedly to write $\phi_k = f_k(1+\xi_k)$ where ξ_k depends only on $x_1, ..., x_k$ (but we omit the dependence) and, by Theorem 18 and the definition of f, $\xi_k = O(S^{-\frac{1}{2}-w\tau})$ for some positive w. Furthermore, let $\delta_k = 0$ if $k = a_i$ or $k = a_i - 1$ and $\delta_k = 1$ otherwise.

These steps yield

$$\Delta_i = \tag{3.61}$$

$$\frac{(a_i-1)[\sum_{D_1,D_0}(\Pi_{x_k\in D_1}f_k(x_k)(1+\xi_k\delta_k)\Pi_{x_k\in D_0}f_k(x_k)-\Pi_{x_k\in D_1}f_k(x_k)\Pi_{x_k\in D_0}f_k(x_k)(1+\xi_k\delta_k))+\epsilon]}{\sum_{D_0}\Pi_{k=1}^{a_i}\phi_k(x_k)\sum_{D_0}\Pi_{k=1}^{a_i}f_k(x_k)}$$

where ϵ is the compensatory term for zeroing out certain terms by inserting the δ_k into equation (3.61), which we can express as $\epsilon = \epsilon_1 + \epsilon_2 - \epsilon_3 - \epsilon_4$ for

$$\epsilon_1 = \sum_{D_1}\xi_{a_i}f_{a_i}(x_{a_i})f_{a_i-1}(x_{a_i-1})\Pi_{k\neq a_i-1,a_i}f_k(x_k)(1+\xi_k)\sum_{D_0}\Pi_{x_k\in D_0}f_k(x_k), \tag{3.62}$$

$$\epsilon_2 = \sum_{D_1}\xi_{a_i-1}f_{a_i-1}(x_{a_i-1})\Pi_{k\neq a_i-1}f_k(x_k)(1+\xi_k)\sum_{D_0}\Pi_{x_k\in D_0}f_k(x_k), \tag{3.63}$$

$$\epsilon_3 = \sum_{D_1}\Pi_{x_k\in D_1}f_k(x_k)\sum_{D_0}\xi_{a_i-1}f_{a_i}(x_{a_i})f_{a_i-1}(x_{a_i-1})\Pi_{k\neq a_i-1,a_i}f_k(x_k)(1+\xi_k), \tag{3.64}$$

$$\epsilon_4 = \sum_{D_1}\Pi_{x_k\in D_1}f_k(x_k)\sum_{D_0}\xi_{a_i}f_{a_i}(x_{a_i})\Pi_{k\neq a_i}f_k(x_k)(1+\xi_k). \tag{3.65}$$

The procedure for identifying that ϵ is indeed a 'higher order term' is exactly identical to the proof in Theorem 9 and hence we ignore the contribution from ϵ.

It follows instantly from equation (3.61) that once we distribute $\Pi f_k(1+\xi_k\delta_k)$ and cancel identical terms, the contribution in the numerator only consists of terms where there is at least one ξ_k in the product. Hence we can define a vector η where each entry η_k either equals 1 or 0 but $\eta \neq \mathbf{0}$ and $\eta_{a_i-1} = \eta_{a_i} = 0$, which yields

$$\Delta_i \leq \sum_{\eta\neq\mathbf{0}}\frac{(a_i-1)[\sum_{D_1,D_0}\Pi_{x_k\in D_1}f_k(x_k)(1+\xi_k\delta_k-\eta_k)\Pi_{x_k\in D_0}f_k(x_k)]}{\sum_{D_0}\Pi_{k=1}^{a_i}\phi_k(x_k)\sum_{D_0}\Pi_{k=1}^{a_i}f_k(x_k)} - . \tag{3.66}$$

$$\sum_{\eta\neq\mathbf{0}}\frac{(a_i-1)[\sum_{D_1,D_0}\Pi_{x_k\in D_1}f_k(x_k)\Pi_{x_k\in D_0}f_k(x_k)(1+\xi_k\delta_k-\eta_k)]}{\sum_{D_0}\Pi_{k=1}^{a_i}\phi_k(x_k)\sum_{D_0}\Pi_{k=1}^{a_i}f_k(x_k)}.$$

Now, for $k = a_i$ or $k = a_i - 1$, $f_k(x_k)(1 + \xi_k \delta_k) = f_k(x_k)$ by definition. For these choices of k in D_0 and D_1, we employ the relationship that $f_k = \mathfrak{i}_k + z_k$ where $\mathfrak{i}_k$ is independent of $x_1, ..., x_{k-1}$ and $z_k = O(S^{-\frac{1}{2}-\tau})f_k$ depends on these values. The z_k represent the new component here relative to Theorem 9, and we need to show that these terms are small enough that they do not make an important contribution. We denote the first $a_i - 2$ variables in D_0 and D_1 as F_0 and F_1 respectively. We now have the following (omitting the dependence on x):

$$\Delta_i \leq \sum_{\eta \neq \mathbf{0}} \frac{(a_i - 1)\sum_{D_1}[\mathfrak{i}_{a_i-1} + z_{a_i-1}]^2 \Pi_{x_k \in F_1} f_k(1 + \xi_k - \eta_k) \sum_{D_0}[\mathfrak{i}_{a_i-1} + z_{a_i-1}][\mathfrak{i}_{a_i} + z_{a_i}]\Pi_{x_k \in F_0} f_k}{\sum_{D_0} \Pi_{k=1}^{a_i} \phi_k \sum_{D_0} \Pi_{k=1}^{a_i} f_k} - \frac{(a_i - 1)\sum_{D_1}[\mathfrak{i}_{a_i-1} + z_{a_i-1}]^2 \Pi_{x_k \in F_1} f_k \sum_{D_0}[\mathfrak{i}_{a_i-1} + z_{a_i-1}][\mathfrak{i}_{a_i} + z_{a_i}]\Pi_{x_k \in F_0} f_k(1 + \xi_k - \eta_k)}{\sum_{D_0} \Pi_{k=1}^{a_i} \phi_k \sum_{D_0} \Pi_{k=1}^{a_i} f_k}. \tag{3.67}$$

With some algebra, we can re-express equation (3.67) as

$$\Delta_i \leq \sum_{\eta \neq \mathbf{0}} \frac{(a_i - 1)\sum_{D_1}[\mathfrak{i}_{a_i-1}]^2 \Pi_{x_k \in F_1} f_k(1 + \xi_k - \eta_k) \sum_{D_0}[\mathfrak{i}_{a_i-1}][\mathfrak{i}_{a_i}]\Pi_{x_k \in F_0} f_k}{\sum_{D_0} \Pi_{k=1}^{a_i} \phi_k \sum_{D_0} \Pi_{k=1}^{a_i} f_k} - \frac{(a_i - 1)\sum_{D_1}[\mathfrak{i}_{a_i-1}]^2 \Pi_{x_k \in F_1} f_k \sum_{D_0}[\mathfrak{i}_{a_i-1}][\mathfrak{i}_{a_i}]\Pi_{x_k \in F_0} f_k(1 + \xi_k - \eta_k)}{\sum_{D_0} \Pi_{k=1}^{a_i} \phi_k \sum_{D_0} \Pi_{k=1}^{a_i} f_k} + \Omega \tag{3.68}$$

where Ω is the compensatory term for omitting the z_k terms (which we define later). Now since $\mathfrak{i}_k$ is independent of $x_1, ..., x_{k-1}$, we can employ the Mean Value Theorem to integrate out the last 2 variables, define D_* as the set of sets of $a_i - 2$ distinct indices, and express equation (3.68) as

$$\Delta_i \leq \tag{3.69}$$

$$\sum_{\eta \neq \mathbf{0}} \frac{(a_i - 1)\sum_{D_*, D_*} \lambda_1 \Pi_{x_k \in D_*} f_k(1 + \xi_k - \eta_k)\lambda_2 \Pi_{x_k \in D_*} f_k - \lambda_1 \Pi_{x_k \in D_*} f_k \lambda_2 \Pi_{x_k \in D_*} f_k(1 + \xi_k - \eta_k)}{\sum_{D_0} \Pi_{k=1}^{a_i} \phi_k \sum_{D_0} \Pi_{k=1}^{a_i} f_k} + \Omega =$$

$$\sum_{\eta \neq \mathbf{0}} \frac{(a_i - 1)\lambda_1 \lambda_2 \sum_{D_*, D_*}[\Pi_{x_k \in D_*} f_k(1 + \xi_k - \eta_k)\Pi_{x_k \in D_*} f_k - \Pi_{x_k \in D_*} f_k \Pi_{x_k \in D_*} f_k(1 + \xi_k - \eta_k)]}{\sum_{D_0} \Pi_{k=1}^{a_i} \phi_k \sum_{D_0} \Pi_{k=1}^{a_i} f_k} + \Omega = \sum_{\eta \neq \mathbf{0}} \Omega.$$

Now we conclude the proof. From equation (3.67) we can write $\Omega = \Omega_+ - \Omega_-$ where Ω_+ includes the positive compensatory terms from the (first part of) equation (3.67) and Ω_- includes the negative compensatory terms from the (second part of) equation 3.67. The following table illustrates all of the subcases that constitute the $\Omega'_{+,i}s$ in equation 3.67. The columns indicate the term in the first summation found in equation (3.67) (Ω_+) and the rows indicate the term in the second summation. If the i,jth entry is a ✓, then this is a relevant subcase, while the case marked with **X** is not.

	$\mathfrak{i}^2_{a_i-1}$	$2\mathfrak{i}_{a_i-1}z_{a_i-1}$	$z^2_{a_i-1}$
$\mathfrak{i}_{a_i-1}\mathfrak{i}_{a_i}$	**X**	✓	✓
$\mathfrak{i}_{a_i-1}z_{a_i}$	✓	✓	✓
$\mathfrak{i}_{a_i}z_{a_i-1}$	✓	✓	✓
$z_{a_i}z_{a_i-1}$	✓	✓	✓

Note that only the first entry in the table is **X**; we already demonstrated through equation (3.68) that the contribution from the the case where both summations contribute $\mathfrak{i}'s$ without any $z's$ is 0. Hence, we have 11 subcases and we can express $\Omega_+ = \sum_{i=1}^{11}\Omega_{+,i}$ and $\Omega_- = \sum_{i=1}^{11}\Omega_{-,i}$.

As we have demonstrated in this work, the z's represent terms that are much smaller than the $\mathfrak{i}'s$. Consequently, *interesting* subcases will involve the fewest number of appearances fo z's as possible. Furthermore, by symmetry the terms corresponding to rows involving the $\mathfrak{i}_{a_i-1}z_{a_i}$ or $\mathfrak{i}_{a_i}z_{a_i-1}$ terms are identical due to symmetry. We therefore only need to examine two subcases, (a) the subcase where the first summation is $2\mathfrak{i}_{a_i-1}z_{a_i-1}$ and the second summation is $\mathfrak{i}_{a_i-1}\mathfrak{i}_{a_i}$, which we define to be $\Omega_{+,1}, \Omega_{-,1}$ and (b) the subcase where the first summation is $\mathfrak{i}^2_{a_i-1}$ and the second summation is $\mathfrak{i}_{a_i-1}z_{a_i}$ which we define to be $\Omega_{+,3}, \Omega_{-,3}$. (We leave the subcase (b) as an exercise to the reader.)

So consider,

$$\Omega_{+,1} = \sum_{\eta\neq\mathbf{0}} \frac{(a_i-1)\sum_{D_1}[2\mathfrak{i}_{a_i-1}z_{a_i-1}]\Pi_{x_k\in F_1}f_k(1+\xi_k-\eta_k)\sum_{D_0}[\mathfrak{i}_{a_i-1}][\mathfrak{i}_{a_i}]\Pi_{x_k\in F_0}f_k}{\sum_{D_0}\Pi_{k=1}^{a_i}\phi_k\sum_{D_0}\Pi_{k=1}^{a_i}f_k}. \quad (3.70)$$

We will briefly consider the case where there is precisely one $\eta_k = 1$. By symmetry we will consider $\eta_1 = 1$ and multiply this quantity by $a_i - 2$ (as there are $a_i - 2$ possible choices to let $\eta_k = 1$; recall that $\eta_{a_i-1} = \eta_{a_i} = 0$), obtaining

$$\Omega^*_{+,1} := \frac{(a_i-2)(a_i-1)\sum_{D_1}[2\mathfrak{i}_{a_i-1}z_{a_i-1}]f_1\xi_1\Pi_{x_k,k\neq 1\in F_1}f_k(1+\xi_k)\sum_{D_0}[\mathfrak{i}_{a_i-1}][\mathfrak{i}_{a_i}]\Pi_{x_k\in F_0}f_k}{\sum_{D_0}\Pi_{k=1}^{a_i}\phi_k\sum_{D_0}\Pi_{k=1}^{a_i}f_k}. \tag{3.71}$$

But by assumption $z_{a_i-1} \leq O(S^{-\frac{1}{2}-\tau})f_{a_i-1}$ and trivially $\mathfrak{i}_{a_i-1} \leq Cf_{a_i-1}$, for some constant C, so we have that

$$\Omega^*_{+,1} \leq \frac{(a_i-2)(a_i-1)\sum_{D_1}[O(S^{-\frac{1}{2}-\tau})f^2_{a_i-1}]f_1\xi_1\Pi_{x_k,k\neq 1\in F_1}f_k(1+\xi_k)\sum_{D_0}[f_{a_i-1}][f_{a_i}]\Pi_{x_k\in F_0}f_k}{\sum_{D_0}\Pi_{k=1}^{a_i}\phi_k\sum_{D_0}\Pi_{k=1}^{a_i}f_k}. \tag{3.72}$$

Now applying the fact that $a_i \leq d_{max} = O(S^{\frac{1}{2}-\tau})$ and $\xi_1 = O(S^{-\frac{1}{2}-w\tau})$,

$$\Omega^*_{+,1} \leq O(S^{-(3+w)\tau})\frac{\sum_{D_1}[2f^2_{a_i-1}]f_1\Pi_{x_k,k\neq 1\in F_1}f_k(1+\xi_k)\sum_{D_0}[f_{a_i-1}][f_{a_i}]\Pi_{x_k\in F_0}f_k}{\sum_{D_0}\Pi_{k=1}^{a_i}\phi_k\sum_{D_0}\Pi_{k=1}^{a_i}f_k}. \tag{3.73}$$

Then employing the technique we used in the proof of Theorem 9 by applying Corollary 9 to reduce the domains D_0,D_1 to D_* we get that

$$\Omega^*_{+,1} \leq O(S^{-(3+w)\tau})O(S^{\frac{7}{2}-\tau})\frac{\sum_{D_*}\Pi_{x_k\in D_*}f_k(1+\xi_k)\sum_{D_*}\Pi_{x_k\in D_*}f_k}{O(S^4)\sum_{D_*}\Pi_{k=1}^{a_i-2}\phi_k\sum_{D_*}\Pi_{k=1}^{a_i-2}f_k} = O(S^{-\frac{1}{2}-4\tau-w\tau}), \tag{3.74}$$

and hence this constribution is a higher order term. Finally, we only considered the case where there is precisely one $\eta_k = 1$. But cases where there are more positive $\eta_k's$ yield even smaller terms! For example, when we had precisely one $\eta_k = 1$, there were at most a_i choices for selecting k, but for each $\eta_k = 1$, we end up multiplying the term f_k by ξ_k (as opposed to $(1+\xi_k)$), and hence if there are m $\eta_k's$ that equal 1, then this results in a contribution bounded by $\Omega_{+,1} \leq \sum_{m=1}^{a_i}(a_i)^{m-1}\xi^{m-1}\Omega^*_{+,1} = \sum_{m=1}^{a_i}O(S^{-w(m-1)\tau-(m-1)\tau})\Omega^*_{+,1} = O(S^{-\frac{1}{2}-4\tau-w\tau})$. Since these contributions are now much smaller than $O(S^{-\frac{1}{2}-2\tau-w\tau})$, we conclude that

$$\Delta_i = O(S^{-\frac{1}{2}-2\tau-w\tau})$$

and from the beginning part of the proof of Theorem 9 conclude that our new approximation $\frac{\|G_{\mathbf{d_{-i}}}\|}{\|G_{\mathbf{d_{-j}}}\|} = g(\mathbf{e}_i, \mathbf{e_j}, \mathbf{d}, \mathbf{a})(1 + O(S^{-\frac{1}{2}-(w+2)\tau}))$.

□

3.7 SOME EXTENSIONS AND COROLLARIES

3.7.1 Maximum of the Degree Sequence

In this section, we briefly mention a generalization of the sparsity assumptions that will also yield a power series expansion for the number of graphs realizing a degree sequence, using the techniques of this paper.

Specifically, up to this point, we have assumed that $d_{max} = O(S^{\frac{1}{2}-\tau})$. Denote $a^{(k)}$ ($b^{(k)}$) as the kth entry in an in-degree (out-degree) sequence derived by sorting $\mathbf{a}$ and $\mathbf{b}$ into non-increasing order. The key observation is that if

$$\max(\sum_{k=1}^{a^{(1)}} b^{(k)}, \sum_{k=1}^{b^{(1)}} a^{(k)}) = O(S^{1-\tau}), \tag{3.75}$$

then the appropriate extensions of Corollary 4 and Lemma 2 still hold and we can derive the corresponding power expansions. Naturally, for example, condition (3.75) holds if $a^{(1)}b^{(1)} = O(S^{1-\tau})$ or if more simply $d_{max} = O(S^{\frac{1}{2}-\tau})$.

Recall the proof of Theorem 2, where we count the number of common neighbors that receive an outward (inward) edge from both an arbitrary node x and a node of bounded degree y. In the worst case scenario, (3.75) gives the number of outgoing edges from all of the neighbors of x. Since there are S edges in the graph, it is intuitive that it would be difficult for x and y, which has bounded degree, to share a common neighbor. This idea forms the foundation for the appropriate extensions of Theorem 2 and Corollaries 3 and 4.

Similarly, with care, we can extend Lemma 2 as follows:

Lemma 6. *Suppose that $f, g : I := \{1, 2, ..., N\} \to [1, \infty)$ and for simplicity let $g(\cdot) \le f(\cdot)$. Let $\{x_1, ..., x_r\}$ be distinct inputs from I that yield the largest r outputs for $f(\cdot)$ and assume* *that* $\sum_{i=1}^{r} f(x_i) = O(S^{1-\tau})$, $\sum_{i=1}^{N} f(i) = O(S)$. *Furthermore, let* k *be an* $O(1)$ *natural number and let* $c_1, \ldots, c_r$ *be a sequence of natural numbers. Then*

$$\sum_{c_1 \neq \ldots \neq c_r}^{N} g(c_1)\Pi_{i=2}^{r} f(c_i) = \sum_{\substack{c_1,..,c_k, \\ c_{k+1} \neq \ldots \neq c_r}}^{N} g(c_1)\Pi_{i=2}^{r} f(c_i)(1 + O(S^{-\tau})).$$

We omit the details of the proofs of these results.

3.7.2 Graphs without Loops and Undirected Graphs

In this section, we discuss how to generalize results from previous sections to directed graphs without loops and to undirected graphs (with and without loops); similar ideas apply to graphs with other sets of prohibited edges besides loops.

3.7.2.1 Directed Graphs without Loops As one may expect, when considering directed graphs without loops, the results regarding asymptotic enumeration for directed graphs with loops carry over as the likelihood that a fixed node has an edge to itself is small. More specifically,

Lemma 7. *Consider a degree sequence* $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{Z}^{N \times 2}$ *where*

$$\mathbf{a} = \{a_1, a_2, 0, \ldots, 0\}, \mathbf{b} = \{\delta, \delta, 2, ..., 2, 1, .., 1, 0, .., 0\} \text{ and } \sum_{i=1}^{N} a_i = \sum_{i=1}^{k} b_i =: S,$$

with δ *either* 0 *or* 1 *and with 2 appearing* q *times in* $\mathbf{b}$. *If* $a_1, a_2 \geq q + \delta$, *then there are* $\binom{a_1+a_2-2\delta-2q}{a_1-\delta-q}$ *directed graphs without loops that realize the above degree sequence.*

The above Lemma allows for the appropriate generalization of Theorem 7, where we construct sets of residual degree sequences X_{k,δ_1,δ_2} for k the number of common neighbors being considered. Here, $\delta_1 = 1$ if node 1 connects to node 2 and is otherwise 0; similarly, $\delta_2 = 1$ if node 2 connects to node 1 and is otherwise 0. Since it should be unlikely asymptotically that either δ_1 or δ_2 is positive, it follows that $\|G_{X_{0,0,0}}\|$ becomes the dominating term and the analysis proceeds as in other cases. (We can make this rigorous by using switching arguments as we did in Section 3.) This will provide us with arbitrarily accurate asymptotics for the ratio of the number of directed graphs (without loops) with degree sequences that are distance 2 apart.

Suppose we knew exactly the number of graphs of one degree sequence that summed to S. Then we could multiply this quantity by a product of (fractional) terms, where each fractional term is a ratio of the number of directed graphs without loops, distance 2 apart. Since from above we have asymptotics for the ratio of the number of graphs distance two apart, we only need one case where we can count the number of graphs of a given degree sequence exactly to derive the general formula. For simplicity, we can consider a degree sequence where all realizations can never have loops. For instance, we can arbitrarily extend our degree sequence to an analogous degree sequence in $\mathbb{Z}^{2N\times 2}$ and we can consider the degree sequence $(\{a_1, ..., a_N, 0, ..., 0\}, \{0, ..., 0, b_{N+1}, ..., b_{2N})\} \in \mathbb{Z}^{2N\times 2}$. If $b_{N+1}, ..., b_{2N}$ consists of only zeros and ones, then Corollary 11 applies and all realizations are directed graphs without loops.

3.7.2.2 Undirected Graphs Practically speaking, the primary difference between manipulating degree sequences for undirected graphs and degree sequences for directed graphs is that for directed graphs, the in-degree sequence can be manipulated without impacting the out-degree sequence. In contrast, with undirected graphs, the corresponding adjacency matrix is symmetric. As such, while the same thematic ideas also follow through to undirected graphs, care should be taken. More specifically, instead of partitioning two rows (or two columns) separately, as we were able to do in the directed case, we must partition the two rows and the two columns together. In the remainder of this section, we will present results in terms of undirected graphs without loops, noting that the techniques carry over to the case where loops are allowed.

For the undirected case, we must consider two distinct cases. In the first case, the two partitioned nodes do not share an edge together. Consequently, if the two nodes have $a_1 + a_2$ edges, then these edges must also show up in the degrees of $a_1 + a_2$ other nodes in the graph.

Lemma 8. *Consider a degree sequence* $\mathbf{a} \in \mathbb{Z}^N$ *for an undirected graph without loops, given by*

$$\mathbf{a} = \{a_1, a_2, 2, ..., 2, 1, ..., 1, 0, ..., 0\}$$

where the first two entires of $\mathbf{a}$ *are followed by* q_2 $2's$ *and* q_1 $1's$ *such that* $q_1 + 2q_2 = a_1 + a_2$*. Then there are* $\binom{a_1+a_2-2q_2}{a_1-q_2}$ *undirected graphs without loops that realize the above degree sequence where nodes 1 and 2 are not neighbors.*

Alternatively, if nodes 1 and 2 are neighbors then they make $a_1 + a_2 - 2$ edges with other nodes in the graph. We need to consider this case as well and the result is analogous to Lemma 8. (If we were considering undirected graphs with loops, we would also need to consider the case where there are self-loops too.) Then as before, we construct sets $X_{k,\delta}$, where k denotes the number of common neighbors of nodes 1 and 2 and δ denotes whether nodes 1 and 2 are linked to each other. Again, $\|G_{X_{0,0}}\|$ dominates and the analogous results hold. To switch from ratios to counts for a particular degree sequence, the following result, which is analogous to Corollary 6, is fundamental.

Lemma 9. *For a degree sequence* $\mathbf{a} \in \mathbb{Z}^N$ *for an undirected graph without loops given by* $\mathbf{a} = \{1, ..., 1, 0, ..., 0\}$ *with* $\sum a_i = S$*,*

$$\|G_{\mathbf{a}}\| = \frac{S!}{2^{\frac{S}{2}}(\frac{S}{2}!)}.$$

Proof. The proof is constructive. Note that S must be even as otherwise there are no graphs. For the first edge, there are $\binom{S}{2}$ possible choices of pairs of nodes to connect. After we decide the initial pair to wire up, then there are $\binom{S-2}{2}$ possible choices to form the second edge. This reasoning implies that there are $\Pi_{k=0}^{\frac{S}{2}-1}\binom{S-2k}{2} = \frac{S!}{2^{S/2}}$ choices for wirings. In this procedure, however, suppose that all of the choices are the same except for the first two steps. In the first step in one example, node 1 wires with node 2 and node 3 wires with node 4. Alternatively in the other example, node 3 wires up with node 4 in the first step, but node 1 wires up with node 2 in the second step. The output is the same graph, but we counted both of these instances as two distinct events (graphs). We readily note that there are $\frac{S}{2}!$ possible ways of wiring up the same graph with this procedure, as there are $\frac{S}{2}$ edge pairs in the graph. □

The technique for counting the number of undirected graphs with loops for the degree sequence $\mathbf{a} = \{1, ..., 1, 0, ..., 0\}$ is similar and left as an exercise to the reader. At this juncture, we note that the idea of prohibiting loops is a special case of prohibiting edges between (two) nodes. Though it is outside the scope of this work, we strongly expect that the ideas can be extended to attain asymptotics for the ratio of the number of graphs of two distinct degree sequences distance 2 apart where we prohibit edges between certain nodes. As long as generating a prohibited edge is unlikely for any node in the graph, the ideas of our work should carry over; that is, the dominating term in the ratio will consist of realizations of graphs where the two partitioned nodes do not share a common neighbor and no prohibited edges appear.

3.7.3 Computing Probabilities

While enumerative asymptotics have been employed for constructing realizations of the Uniform Model [8], we can also construct probabilities for the likelihood that two nodes share an edge and compare these probabilities to other random graph models. Suppose node 1 has degree a_1. Then the probability node 1 receives an outward edge from node 2, which we denote as $Pr(2 \to 1)$, would be

$$Pr(2 \to 1) = \frac{\frac{1}{(a_1-1)!} \sum_{x_1=2 \neq x_2 \ldots \neq x_{a_1-1}}^{N} \|G_{(\mathbf{a}-a_1\mathbf{e}_1, \mathbf{b}-\sum_{i=1}^{a_1} \mathbf{e}_{x_i})}\|}{\frac{1}{a_1!} \sum_{x_1 \neq \ldots \neq x_{a_1}}^{N} \|G_{(\mathbf{a}-a_1\mathbf{e}_1, \mathbf{b}-\sum_{i=1}^{a_1} \mathbf{e}_{x_i})}\|} \tag{3.76}$$

To motivate (3.76), suppose we are given our degree sequence and decide to wire up the edges corresponding to node 1. Our numerator simply counts the number of graphs where we have wired up node 1 with a_1 other nodes (as node 1 has in-degree a_1) and one of those nodes must be node 2. Similarly, our denominator counts the number of graphs where we have wired up node 1 with a_1 other nodes. Using the techniques of Chapter 3, we can simplify this further by considering ratios of the form, $\frac{\|G_{(\mathbf{a},\mathbf{b}-\mathbf{e}_*)}\|}{\|G_{(\mathbf{a},\mathbf{b}-\mathbf{e}_{**})}\|}$. We can then express (3.76) as a sum of products of ratios of the number of graphs from two slightly different degree sequences,

$$Pr(2 \to 1) = \frac{\frac{1}{(a_1-1)!} \sum_{\substack{x_1=2 \neq \\ x_2 \neq \ldots \neq x_{a_1}}}^{N} \prod_{j=1}^{a_1} \frac{\|G_{\mathbf{a}-a_1\mathbf{e}_1, \mathbf{b}-\sum_{k=1}^{a_1-j} \mathbf{e}_{v_k} - \sum_{k=1}^{j} \mathbf{e}_{x_j}}\|}{\|G_{\mathbf{a}-a_1\mathbf{e}_1, \mathbf{b}-\sum_{k=1}^{a_1-j+1} \mathbf{e}_{v_k} - \sum_{k=1}^{j-1} \mathbf{e}_{x_j}}\|}}{\frac{1}{a_1!} \sum_{x_1 \neq \ldots \neq x_{a_1}}^{N} \prod_{j=1}^{a_1} \frac{\|G_{\mathbf{a}-a_1\mathbf{e}_1, \mathbf{b}-\sum_{k=1}^{a_1-j} \mathbf{e}_{v_k} - \sum_{k=1}^{j} \mathbf{e}_{x_j}}\|}{\|G_{\mathbf{a}-a_1\mathbf{e}_1, \mathbf{b}-\sum_{k=1}^{a_1-j+1} \mathbf{e}_{v_k} - \sum_{k=1}^{j-1} \mathbf{e}_{x_j}}\|}} \tag{3.77}$$

Using our results from Chapter 3: Corollary 9, Theorem 16 and Theorem 18 to reduce (3.77) to

$$Pr(2 \to 1) \approx \frac{a_1 b_2}{S}(1 + O(S^{-2\tau})) \tag{3.78}$$

and hence we conclude that asymptotically, the probability that two nodes share an edge under the uniform random graph converge to the probabilities in the Chung-Lu random graph model! (Recall that we defined the Chung-Lu random graph model in the introduction). Note that we attain convergence with a factor of $(1 + O(S^{-2\tau}))$ where $\max_i a_i = O(S^{\frac{1}{2}-\tau})$ and $\max_i b_i = O(S^{\frac{1}{2}-\tau})$. Note that since for many degree sequences τ may be small, the Uniform random graph model could differ dramatically from the Chung-Lu random graph model.

4.0 CONCENTRATION RESULTS REGARDING THE SPECTRAL RADIUS FOR RANDOM DIRECTED CHUNG-LU GRAPHS

In this chapter, we seek mathematically rigorous asymptotic results for the dominating eigenvalue of an adjacency matrix under different random directed graph models for the Chung-Lu random graph model as well as a generalization that enables community structure within the graph. We also attain bounds for how much the dominating eigenvalue can deviate from the asymptotics for a realization of a graph with fixed N nodes. Finally, we apply our results in studying the stability of a epidemic (SIS) contact network and the stability of synchrony in the Kuramoto model.

Random graph models are useful for constructing families of graphs where all graphs in the collection possess a similar property. More specifically, the dominating eigenvalue of the adjacency matrix in biological neural networks, boolean genetic networks and epidemiological networks plays an important role in the stability of solutions to the corresponding dynamical systems and stochastic processes that occur on these networks. To construct meaningful statements regarding the aforementioned networks, we also want to utilize a random graph model that shares characteristics that matches empirically observed real world networks. One such property is degree heterogeneity, that is, empirically observed networks have nodes that possess different numbers of neighbors throughout the network. While we focus on directed graphs, we can extend many of the proof techniques in this work to undirected graphs as well. At this juncture, we introduce the definition of the Chung-Lu random graph model, which allows for the generation of networks with degree heterogeneity.

Definition 6 (Directed Chung-Lu Model)**.** *Suppose we have a vector of weights* $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{Z}^{N\times 2}$ *(an expected bi-degree sequence), where* $S = \sum_{a_i \in \mathbf{a}} a_i = \sum_{b_i \in \mathbf{b}} b_i$ *and* $\max_{i,j} a_i b_j \leq S$*.*

We construct an edge from node i to node j with an independent Bernoulli random variable of probability $p_{ij} = \frac{b_i a_j}{S}$

Heuristic arguments have suggested the following result [73],

Conjecture 1. *Given a sequence of realizations of the Directed Chung-Lu model where we consider increasingly large networks with N nodes. Denote the absolute value of the dominating eigenvalue of A by $\rho(A)$. Then almost surely,*

$$\lim_{N \to \infty} \frac{\rho(A)}{\frac{\mathbf{a} \cdot \mathbf{b}}{S}} = 1$$

When constructing bounds on the dominating maximum eigenvalue of an asymmetric (directed) matrix, we do not have quite as many tools to construct concentration results compared to the symmetric (undirected) case. Our main theoretical tool will be the following lemma, but first we introduce the following definition for clarity.

Definition 7. *We define the spectral radius of a matrix $A \in \mathbb{R}^{N \times N}$ to be the absolute value of the maximal magnitude eigenvalue. More precisely,*

$$\rho(A) = max\{|\lambda| : Ax = \lambda x \text{ and } x \neq \mathbf{0}\}$$

Lemma 10. *Let $A \in \mathbb{R}^{N \times N}$ be an entrywise non-negative matrix. For simplicity, suppose that the eigenvalues of A are real.*

Define λ_{max} as the maximum of the eigenvalues of A. It then follows that $\rho(A) = \lambda_{max}$.

Furthermore, we have for every positive integer r,

$$\frac{trace(A^r)}{N} \leq \rho(A^r) = \rho(A)^r \leq \mathbf{1}^T A^r \mathbf{1} \tag{4.1}$$

where $\mathbf{1}$ is the vector of one's.

For completeness we provide the following proof.

Proof. First, note that for an entry-wise non-negative matrix A, for any positive integer k, A^k, will also be an entry-wise non-negative matrix. Now suppose that λ_{max}, the maximum of the eigenvalues of A was in fact negative. Then since the trace of a matrix is the sum of the eigenvalues, it follows that for some choice of m, $trace(A^m) < 0$, but this is a contradiction since A^m is entry-wise non-negative. As such we conclude that $\lambda_{max} = \rho(A)$.

The lower bound to $\rho(A^r)$ in (4.1) follows directly from the fact that $\rho(A^r)$, equals the largest nonnegative eigenvalue of A^r and hence is also the maximum of the eigenvalues of A^r and $\frac{trace(A^r)}{N}$ is the average of the eigenvalues of A^r.

Next, the equality in (4.1) follows from the fact that if $\lambda, \mathbf{x}$ is an eigenvalue, eigenvector pair of A, then $\lambda^r, \mathbf{x}$ is an eigenvalue,eigenvector pair of A^r where r is a positive integer. Note that from the Jordan Cannonical Form that all eigenvalues of A^r are of the form λ^r where λ is an eigenvalue of A. It then follows that $0 \leq \lambda_{max} = \rho(A) \implies \rho(A)^r = \lambda_{max}^r = \rho(A^r)$.

And finally, to prove the upperbound on $\rho(A^r)$, let $\mathbf{v}$ satisfy $A\mathbf{v} = \lambda_{max}\mathbf{v}$ where we normalize $\mathbf{v}$ such that $\mathbf{v}^T\mathbf{v} = 1$. Define $|\mathbf{v}|$ to be the vector consisting of the absolute value of the entries of $\mathbf{v}$. Since A is entry-wise non-negative, it follows that

$$\mathbf{v}^T A^r \mathbf{v} = \rho(A)^r \leq |\mathbf{v}|^T A^r |\mathbf{v}|.$$

Another application of the fact that A is entry-wise non-negative yields

$$|\mathbf{v}|^T A^r |\mathbf{v}| \leq \mathbf{1}^T A^r \mathbf{1},$$

as $\mathbf{v}^T\mathbf{v} = 1$ implies that $|\mathbf{v}| \leq \mathbf{1}$ entry-wise. □

We are primarily interested in applications of Lemma 10 in studying the spectral radius of an adjacency matrix for random graphs. In this case, if A corresponds to a directed graph, the quantities that bound $\rho(A)$, $\mathbf{1}^T A^r \mathbf{1}$ and $trace(A^r)/N$ both have combinatorial interpretations. More specifically, $trace(A^r) = \sum_{j=1}^{N} \mathbf{e}_j^T A^r \mathbf{e}_j$ (where $\mathbf{e}_j$ is the standard unit vector) is the number of cycles of length r. Similarly,$\mathbf{1}^T A^r \mathbf{1}$ is the number of paths of length r.

To see this consider $\mathbf{1}^T A\mathbf{e}_j$. Now, $A\mathbf{e}_j$ is a vector whose kth entry is 1 if there is an edge from node j to node k. Consequently, $\mathbf{1}^T A\mathbf{e}_j$ is the number of paths starting at node j of length 1 (and analogously $\mathbf{1}^T A\mathbf{1}$ is the number of paths of length 1). One can then proceed inductively, to show that $\mathbf{1}^T A^2\mathbf{e}_j$ is the number of paths starting at node j of length 2 and so on.

We now provide a sketch of our proof technique to prove a version of Conjecture 1. For a suitably chosen r, we will show that with high probability

$$\rho(A)^r \leq \mathbf{1}^T A^r \mathbf{1} \leq (\frac{\mathbf{a}\cdot\mathbf{b}}{S})^r \alpha \tag{4.2}$$

where α depends on the expected degree sequence. We will show that for a **suitable** sequence of expected degree sequences (where we vary r with the expected degree sequence), α will have the property that $\lim_{r\to\infty} \alpha(r)^{\frac{1}{r}} = 1$. Hence it will follow from monotonicity by taking the rth root of both sides of equation (1), that with high probability,

$$\rho(A) \leq \frac{\mathbf{a}\cdot\mathbf{b}}{S}\alpha^{\frac{1}{r}}$$

Perhaps less obvious, we will also show a similar relationship holds for involving the $trace(A^r)$. A conceptually convenient way of seeing this in a *special case* is to note that $\frac{1}{N^2}\mathbf{1}^T A^r \mathbf{1} = \frac{1}{N^2}\sum_{i,j}\mathbf{e}_i^T A^r \mathbf{e}_j \approx \frac{1}{N}\sum_i \mathbf{e}_i^T A^r \mathbf{e}_i = \frac{trace(A^r)}{N} \leq \rho(A)^r$ where the approximation holds for many expected degree sequences.

More succinctly, we will argue that $\frac{1}{N^2}(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^r \leq \frac{1}{N^2}\mathbf{1}^T A^r \mathbf{1} \leq \rho(A)^r \leq \mathbf{1}^T A^r \mathbf{1}$ and if we choose r sufficiently large, say $r = O((logN)^2)$, we have that $(\frac{1}{N^2})^{\frac{1}{r}} \to 1$ and we will be able to conclude that asymptotically

$$\frac{\rho(A)}{\frac{\mathbf{a}\cdot\mathbf{b}}{S}} \to 1.$$

4.1 SPECTRAL CONCENTRATION BOUNDS, $\frac{\mathbf{a}\cdot\mathbf{b}}{S} \to \infty$

Keeping Lemma 10 in mind, we initiate our discussion on bounding the spectral radius for the Chung-Lu model by bounding the expected number of paths of length r. In order to identify *suitable* degree sequences, we will consider two cases seperately, where either $\frac{\mathbf{a}\cdot\mathbf{b}}{S} \to \infty$ (this section) or $p_{max} = \max_{i,j} p_{ij} \to 0$ (next section). Though the results below hold in considerable generality, many of the initially stated results will only be particularly useful for the case where $\frac{\mathbf{a}\cdot\mathbf{b}}{S} \to \infty$. We start by identifying a lower bound on the expectation of $\mathbf{e}_j^T A^r \mathbf{e}_i$ as our first step for constructing spectral radius concentration results.

Lemma 11. *Consider a realization of the Directed Chung-Lu random graph model with expected degree sequence* $\mathbf{d} = (\mathbf{a}, \mathbf{b})$*, where* $\sum a_i = \sum b_i = S$*. Then the expected number of paths from node* i *of length* r *is bounded below by*

$$b_i[\frac{\mathbf{a}\cdot\mathbf{b}}{S}]^{r-1}$$

Furthermore, the expected number of paths of length r *is bounded below by*

$$S[\frac{\mathbf{a}\cdot\mathbf{b}}{S}]^{r-1}$$

Proof. First consider the expected number of paths from node y of length r,

$$\sum_{i_1,\ldots,i_r} Prob[Path(y \to i_1 \to \ldots \to i_r)]$$

Now if no edge repeats in the path (and the probability of each edge is independent), then

$$Prob[Path(y \to i_1 \to \ldots \to i_r)] = p_{yi_1}\Pi_{k=1}^{r-1} p_{i_k i_{k+1}}$$

If however an edge does repeat in the path, since each $0 \le p_{ij} \le 1$ for all i, j, it follows that for such a path

$$Prob[Path(y \to i_1 \to \ldots \to i_r)] \ge p_{yi_1}\Pi_{k=1}^{r-1} p_{i_k i_{k+1}}$$

Consequently, we conclude that the

$$\sum_{i_1,\ldots,i_r} Prob[Path(y \to i_1 \to \ldots \to i_r)] \ge \sum_{i_1,\ldots,i_r} p_{yi_1}\Pi_{k=1}^{r-1} p_{i_k i_{k+1}}$$

Let a_x, b_x denote the expected in-degree/out-degree of node x. Now applying Chung-Lu, we simplify

$$\sum_{i_1,...,i_r} p_{yi_1}\Pi_{k=1}^{r-1}p_{i_k i_{k+1}} = \sum_{i_1,...,i_r} \frac{b_y a_{i_1}}{S}\Pi_{k=1}^{r-1}\frac{b_{i_k}a_{i_{k+1}}}{S} = \sum_{i_1,...,i_r} \frac{b_y a_{i_r}}{S}\Pi_{k=1}^{r-1}\frac{b_{i_k}a_{i_k}}{S}$$

where the last equality follows from rearranging the terms. Now by taking expectations we get that

$$\sum_{i_1,...,i_r} Prob[Path(y \to i_1 \to ... \to i_r)] \geq b_y(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1} \tag{4.3}$$

It then follows that we can construct a lower bound for the expected number of paths of length r by using (4.3), the lower bound of the expected number of paths of length r starting from node y and summing over all possible initial node choices. Since $\sum_i b_i(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1} = S(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1}$, the proof is complete.

□

By counting the number of times repeating edges appear in a given path, the proof of Lemma 11 gives us an immediate corollary that will be helpful in constructing an upper-bound.

Corollary 13. *Consider a realization of the Directed Chung-Lu random graph model with expected degree sequence $\mathbf{d} = (\mathbf{a}, \mathbf{b})$, where $\sum a_i = \sum b_i = S$. The expected number of paths starting at node i of length r where there are no repeating edges is bounded above by $b_i(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1}$. Furthermore, consider the set of all paths P_k where we require the kth edge to be a repeating edge and no other edges can be repeating. Then for a fixed $k < r$, the expected number of paths starting at node i of length r and in P_k are bounded above by $p_{max}b_i(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-3}$, where $p_{max} = \frac{a_{max}b_{max}}{S}$.*

Now we must construct an upper bound for the expected number of paths of length r from a given node y.

Theorem 19. *Denote $p_{max} = \max_{i,j} p_{ij}$. Assume $\frac{S}{\mathbf{a}\cdot\mathbf{b}} < \frac{1}{2}$ and that $r < \frac{\mathbf{a}\cdot\mathbf{b}}{S}$, then we have the following upperbound for the expected number of paths from any node y of length r,*

$$\sum_{i_1,...,i_r} Prob[Path(y \to i_1 \to ... \to i_r)] \leq b_y 2(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1}exp(p_{max}\frac{r^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2}{1 - r\frac{S}{\mathbf{a}\cdot\mathbf{b}}}).$$

Proof. As demonstrated in Lemma 11 when evaluating the likelihood that a particular path exists, it is helpful to identify the edges that repeat multiple times throughout the path. We want a method that identifies 'repeated' edges that will help simplify our calculations.

Consequently, we establish the following rules for determining which edge is a repeating edge.

- The first edge is *never* a repeating edge.
- The last edge can only be a repeating edge if it is identical to the first edge.
- The second edge is a repeating edge if it is identical to the first or last edge.
- And then we proceed inductively, the third edge is a repeating edge if it equals the first, last or second edge, etc.

Now we condition on the possibility that the last edge could be a repeat edge. If the last edge is indeed a repeat edge, then we also know that the last edge has to equal the first edge. Furthermore, we know that the first edge starts at node y, similarly the last edge must also start at node y. We can rewrite this path of length r as a cycle starting and ending at node y (of length $r-1$) where we append this repeated last edge to the cycle. Since this last edge must be identical to the first edge, we conclude that the expected number of paths of length r starting at node y (where the last edge is determined by the first edge in the cycle) equals the expected number of cycles of length $r-1$ starting and ending at node y.

Now define $P_r(y)$ to be the number of paths of length r starting at node y and $C_r(y)$ to be the number of cycles of length r starting at node y. Furthermore, denote $P_r^L(y)$ to be the number of paths of length r where the last edge does not repeat. We have the following decomposition.

$$E(P_r(y)) = E(P_r^L(y)) + E(C_{r-1}(y)) \leq E(P_r^L(y)) + E(P_{r-1}(y)). \tag{4.4}$$

Applying inequality (4.4) over again, we have that $E(P_r(y)) \leq E(P_r^L(y)) + E(P_{r-1}^L(y)) +$

$E(P_{r-2}(y))$. Repeating this trick inductively will yield that

$$E(P_r(y)) \leq \sum_{m=2}^{r} E(P_m^L(y)) + E(P_1(y)) = \sum_{m=1}^{r} E(P_m^L(y)), \tag{4.5}$$

where the last equality follows from the fact that a path of length 1 can never have a repeating edge.

By inequality (4.5), to construct a meaningful bound for $E(P_r(y))$, it will suffice to consider a bound for $E(P_r^L(y))$, **where we are only interested in cases where the last edge (and the first edge) cannot be a repeated edge.** To construct this desired bound, it is helpful to consider some 'simple' cases.

- First, if there are no repeating edges, it follows from Corollary 13 that the expected number of paths that satisfy this property is bounded above by $b_y(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1}$.
- Now suppose that there are repeating edges. For each identical repeating edge, denote the location of all such edges by the set Z (recall that the 'first time' an edge appears, that edge is not considered a repeating edge). Then it follows that the contribution from such paths is,

$$E(P_Z) = \sum_{\substack{i_1,\ldots,i_{r+1} \\ (\beta_k,\beta_{k+1})\in Z}} \frac{b_y a_{i_1}}{S} \Pi_{k=1}^{r} \frac{b_{i_k} a_{i_{k+1}}}{S} \Pi_{(\beta_k,\beta_{k+1})\in Z} \frac{S}{b_{\beta_k} a_{\beta_{k+1}}}$$

- Case 1: Z consists of a unique repeated edge. Then we have that

$$E(P_Z) \leq b_y \frac{a_{max} b_{max}}{S} (\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-3} \tag{4.6}$$

- Case 2: Z consists of 2 edges and both edges are next to each other in the proposed path, then

$$E(P_Z) \leq b_y \frac{a_{max} b_{max}}{S} (\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-4}$$

- Case K: Z consists of K edges and all K edges are next to each other in the proposed path, then

$$E(P_Z) \leq b_y \frac{a_{max} b_{max}}{S} (\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-2-K}$$

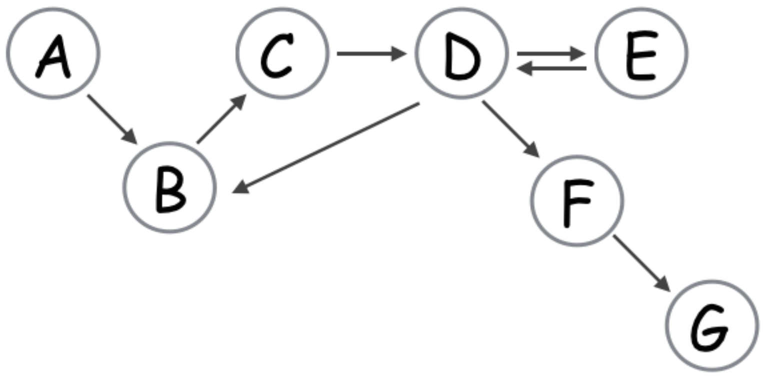

Figure 4.1.1: A Sample Path Illustrating an Example of a Repeating Node Block.

To clarify our proof strategy, we consider the following example, the path taken in Figure 4.1.1: $A \to B \to C \to D \to B \to C \to D \to E \to D \to F \to G$. Each node in the path has either never been visited before or has been visited before. Furthermore, if the node has been visited before, it is part of a block of previously visited nodes of some prescribed length. In the path above, we initially have four nodes that have never been visited before (A-D) then a block of three previously visited nodes (B-D), followed by a new node E, another 'block' that we already visited (node D) and then two more nodes that we never visited before. Recall that we defined $p_{max} = a_{max}b_{max}/S$. We claim that the following expression is an upperbound for the expected number of paths of length r starting from node y where the last edge does not repeat, $P_r^L(y)$.

$$E(P_r^L(y)) \leq b_y \sum_{\substack{k_1+\sum_{i=2}^{r}(i-1)k_i=r \\ k_1\geq 1}} \binom{\sum k_i}{k_1, \ldots, k_r} [k_1^{r-k_1}] (\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{k_1-1} \Pi_{i\geq 2}(p_{max}[\frac{S}{\mathbf{a}\cdot\mathbf{b}}])^{k_i} \tag{4.7}$$

The k_i variables under the summation in equation (4.7) considers the different ways of decomposing a path of length r into repeating node blocks of different sizes. We will justify the above claim, term by term. First we explain the appearance of $\binom{\sum k_i}{k_1,\ldots,k_r}$.

Let k_i denote the number of repeated node blocks of length $i-1$ where k_1 refers to the number of new nodes that has never been visited before (excluding the first node). Then consider all paths of length r with k_i blocks of length $i-1$. Then there are at most

$\binom{\sum k_i}{k_1,\ldots,k_r}$ 'block' rearrangements. Furthermore we also have the constraint that the $k_i \geq 0$ and $k_1 + \sum_{i=2}^{r}(i-1)k_i = r$ as the path by definition has length r, must only contain $r+1$ nodes and by construction k_1 does not include the presence of the first node.

Next, suppose that we have designated the locations (or times) in the path of length r that contain previously visited nodes. For example, in Figure 4.1.1 we fixed the $5, 6, 7$, and 9 places in the path to consist of previously visited nodes. Now each of these repeated nodes must correspond to a node that has never been visited before. In the example, $k_1 = 7$ and $r - k_1 = 4$ as the path in the example has 11 nodes. As such there are at most

$$k_1^{r-k_1}$$

such rearrangements. That is, for each of the $r - k_1$ places (locations) that consist of previously visited nodes, the node visited in each such location must match one of the k_1 distinct nodes in the path.

Suppose for $i \geq 2$, there are k_i blocks of length $i-1$. Then we have the following upperbound for the expected number of paths with the given 'block' structure constraints, $b_y(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{k_1-1}\Pi_{i\geq 2}(p_{max}[\frac{S}{\mathbf{a}\cdot\mathbf{b}}])^{k_i}$. Recall Case 1, equation (4.6), where $k_1 = r-1, k_2 = 1$, all other $k_i = 0$ and the bound for the expected number of paths with only one revisited node is $b_y p_{max}(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-3}$. It can be shown that this formula is consistent with the other aforementioned cases and is indeed the expected number of paths with the given 'block' structure constraints.

But we can trivially bound equation (4.7) by,

$$E(P_r^L(y)) \leq b_y \sum_{\substack{k_1+\sum_{i=2}^{r}(i-1)k_i=r \\ k_1\geq 1}} \binom{\sum k_i}{k_1, \ldots, k_r} (r)^{r-k_1} (\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{k_1-1} \Pi_{i\geq 2} (\frac{a_{max}b_{max}}{S}[\frac{S}{\mathbf{a}\cdot\mathbf{b}}]^1)^{k_i} \leq$$

$$b_y \sum_{k_1+\sum_{i=2}^{r}(i-1)k_i=r} \frac{(\sum_{i=1}^{r} k_i)!}{k_1!} (r)^{r-k_1} (\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{k_1-1} \Pi_{i\geq 2}^{r} \frac{(\frac{a_{max}b_{max}}{S}[\frac{S}{\mathbf{a}\cdot\mathbf{b}}]^1)^{k_i}}{k_i!} \leq \tag{4.8}$$

$$b_y \sum_{k_1+\sum_{i=2}^{r}(i-1)k_i=r} (\sum_{i=1}^{r} k_i)^{\sum_{i=2}^{r} k_i} (r)^{r-k_1} (\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{k_1-1} \Pi_{i\geq 2}^{r} \frac{(\frac{a_{max}b_{max}}{S}[\frac{S}{\mathbf{a}\cdot\mathbf{b}}]^1)^{k_i}}{k_i!} \leq \tag{4.9}$$

But we can write k_1 in terms of the other k_i and r and bound $\sum_{i=1}^{r} k_i$ by r yielding the upperbound,

$$b_y \sum_{k_1+\sum_{i=2}^{r}(i-1)k_i=r} r^{\sum_{i=2}^{r} k_i}(r)^{\sum_{i=2}^{r}(i-1)k_i}(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{k_1-1}\Pi_{i\geq 2}^{r}\frac{(\frac{a_{max}b_{max}}{S}[\frac{S}{\mathbf{a}\cdot\mathbf{b}}]^1)^{k_i}}{k_i!} \leq \tag{4.10}$$

$$b_y \sum_{k_1+\sum_{i=2}^{r}(i-1)k_i=r} r^{\sum_{i=2}^{r} ik_i}(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-\sum_{i=2}^{r}(i-1)k_i-1}\Pi_{i\geq 2}^{r}\frac{(\frac{a_{max}b_{max}}{S}[\frac{S}{\mathbf{a}\cdot\mathbf{b}}]^1)^{k_i}}{k_i!} \leq \tag{4.11}$$

$$b_y(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1} \sum_{k_1+\sum_{i=2}^{r}(i-1)k_i=r} \Pi_{i\geq 2}^{r}\frac{(\frac{a_{max}b_{max}}{S}[r\frac{S}{\mathbf{a}\cdot\mathbf{b}}]^i)^{k_i}}{k_i!} \tag{4.12}$$

And now to achieve an upperbound we can instead allow the $k_i's$ (ignoring k_1) to vary independently resulting in,

$$b_y(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1} \sum_{k_2=0,\ldots,k_r=0}^{\infty} \Pi_{i\geq 2}^{r}\frac{(\frac{a_{max}b_{max}}{S}[r\frac{S}{\mathbf{a}\cdot\mathbf{b}}]^i)^{k_i}}{k_i!} \tag{4.13}$$

Then by identifying the Taylor expansion of $exp(x)$, we get that (4.13) in fact equals,

$$b_y(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1}\Pi_{i=2}^{r}exp(p_{max}r^i(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^i) = b_y(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1}exp(\sum_{i=2}^{r} p_{max}r^i(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^i) \leq \tag{4.14}$$

$$b_y(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1}exp(p_{max}\frac{r^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2}{1-r\frac{S}{\mathbf{a}\cdot\mathbf{b}}}) \tag{4.15}$$

Now from (4.15) and (4.5) we conclude that

$$E(P_r(y)) \leq \sum_{m=1}^{r} b_y(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{m-1}exp(p_{max}\frac{m^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2}{1-m\frac{S}{\mathbf{a}\cdot\mathbf{b}}}) \leq$$

$$b_y(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1}exp(p_{max}\frac{r^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2}{1-r\frac{S}{\mathbf{a}\cdot\mathbf{b}}})\sum_{m=1}^{r}(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^{r-m} \leq$$

$$b_y(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1}exp(p_{max}\frac{r^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2}{1-r\frac{S}{\mathbf{a}\cdot\mathbf{b}}})\frac{1}{1-\frac{S}{\mathbf{a}\cdot\mathbf{b}}} \leq$$

$$2b_y(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1}exp(p_{max}\frac{r^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2}{1-r\frac{S}{\mathbf{a}\cdot\mathbf{b}}})$$

where the last inequality follows from the assumption that $\frac{\mathbf{a}\cdot\mathbf{b}}{S} > 2$.

□

Now that we have results regarding the expectation of the number of paths of length r , we now seek concentration results regarding the dominating eigenvalue of the adjacency matrix.

Theorem 20. *Denote A as a realization of a random Chung-Lu graph with expected degree sequence $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{Z}^{N\times 2}$. Furthermore let $p_{max} = \frac{a_{max}b_{max}}{S}$ and $S = \sum_i a_i = \sum_i b_i$. Then for every $\epsilon \in (0,1)$ there exists an N_1, N_2 such that if $(logN)^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2 < \frac{1}{N_1}$, $N > N_2$ and $(logN)^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}}) < \frac{1}{2}$, then*

$$Prob(\rho(A) \leq (1+\epsilon)\frac{\mathbf{a}\cdot\mathbf{b}}{S}) \geq 1-\epsilon$$

Proof. First, suppose that

$$log(N) > \frac{30}{\epsilon} \text{ and } (logN)^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2 < \frac{\epsilon}{100} \tag{4.16}$$

where we assume without loss of generality that $\epsilon < 1$.

Consider the expected number of paths, P_r, of length r . By Theorem 19, we have that for $r = (logN)^2$,

$$E(P_r) \leq 2S(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1}exp(p_{max}\frac{r^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2}{1-r\frac{S}{\mathbf{a}\cdot\mathbf{b}}}).$$

By Markov's Inequality we have that,

$$Prob(P_r > Z) \leq \frac{E(P_r)}{Z}.$$

Now if $Z = N * E(P_r)$, we have that,

$$Prob(P_r > Z) \leq \frac{1}{N}.$$

It then follows that with probability at least $1 - \frac{1}{N}$ from Lemma 10,

$$\rho(A)^r = \rho(A^r) \leq 2NS(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^r exp(p_{max}\frac{r^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2}{1-r\frac{S}{\mathbf{a}\cdot\mathbf{b}}}). \tag{4.17}$$

Consequently, with probability at least $1 - \frac{1}{N}$,

$$\rho(A) \leq (2NS)^{\frac{1}{r}}\frac{\mathbf{a}\cdot\mathbf{b}}{S}exp(p_{max}\frac{r(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2}{1-r\frac{S}{\mathbf{a}\cdot\mathbf{b}}}) \tag{4.18}$$

But recall that we chose $r = (logN)^2$ so we have that (4.18) simplifies to,

$$\rho(A) \le (2NS)^{\frac{1}{(logN)^2}} \frac{\mathbf{a}\cdot\mathbf{b}}{S} exp(p_{max} \frac{(logN)^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2}{1-(logN)^2\frac{S}{\mathbf{a}\cdot\mathbf{b}}}). \tag{4.19}$$

We first consider the term from (4.19),

$$exp(p_{max} \frac{(logN)^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2}{1-(logN)^2\frac{S}{\mathbf{a}\cdot\mathbf{b}}}). \tag{4.20}$$

By assumption (4.16), we have that (4.20) is bounded above by

$$exp(\frac{2\epsilon}{100}) \le (1+\frac{6\epsilon}{100}), \tag{4.21}$$

where we also used the fact from (4.16) that $(logN)^2\frac{S}{\mathbf{a}\cdot\mathbf{b}} < \frac{1}{2}$ and that $exp(x) \le 1+3x$ for $x < 1$ (since $\epsilon < 1$).

Now we consider the other coefficient from (4.19),

$$(2NS)^{\frac{1}{(logN)^2}} \le (2N^3)^{\frac{1}{(logN)^2}}, \tag{4.22}$$

where the right hand side comes from the fact that the total number of edges must be bounded above by N^2.

We consider the log of the right hand side of (4.22).

$$log(2N^3)^{\frac{1}{(logN)^2}} = \frac{3log(N)+log(2)}{(logN)^2} \le \tag{4.23}$$

$$\frac{4log(N)}{(logN)^2} = \frac{4}{log(N)} < \frac{4\epsilon}{30}. \tag{4.24}$$

So we conclude that (4.22) is bounded above by

$$exp(\frac{4\epsilon}{30}) \le (1+\frac{12\epsilon}{30}). \tag{4.25}$$

since for $x < 1$, $exp(x) \le 1+3x$.

Using (4.21) and (4.25) to bound (4.19), we get that,

$$\rho(A) \le \frac{\mathbf{a}\cdot\mathbf{b}}{S}(1+\frac{12\epsilon}{30})(1+\frac{6\epsilon}{100}) \le \frac{\mathbf{a}\cdot\mathbf{b}}{S}(1+\frac{18\epsilon}{30}+\frac{72\epsilon^2}{3000}) \le \frac{\mathbf{a}\cdot\mathbf{b}}{S}(1+\epsilon) \tag{4.26}$$

for $\epsilon < 1$.

Note that this bound is valid for at least probability $1 - \frac{1}{N}$ where $1 - \epsilon \leq 1 - \frac{\epsilon}{30} \leq 1 - \frac{1}{logN} \leq 1 - \frac{1}{N}$ and the proof is complete.

□

Remark 4. *Under additional assumptions in Theorem 20 regarding how fast* $\frac{\mathbf{a}\cdot\mathbf{b}}{S} \to \infty$*, we can weaken the constraint that* $N > exp(\frac{30}{\epsilon})$ $(\epsilon = O(\frac{1}{log(N)}))$ *to say that with high probability the maximum eigenvalue does not exceed* $\frac{\mathbf{a}\cdot\mathbf{b}}{S}$ *by too much. In particular if* $(\frac{\mathbf{a}\cdot\mathbf{b}}{S}) = O(N)$*, we can choose* $r = O(N)$ *and we will get an analogus theorem to that above where* $\epsilon = O(\frac{log(N)}{N})$*.*

To prove the lower bound, we want to evaluate $\frac{trace(A^r)}{N}$, the average of the number of cycles of length r in the network.

Corollary 14. *Given an expected bidegree sequence* $\mathbf{d}$*, for a given realization* A *it follows that for* $2 < r < \frac{\mathbf{a}\cdot\mathbf{b}}{S}$*,*

$$(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^r \leq E(trace(A^r)) \leq 2(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^r exp(\frac{r^2 p_{max}(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2}{1 - r\frac{S}{\mathbf{a}\cdot\mathbf{b}}}). \quad (4.27)$$

Proof. This Corollary is essentially an extension of Theorem 19, where Theorem 19 provides an upper bound for the expected number of paths of length r, Corollary 14 is a statement about the upper bound for the expected number of cycles of length r. To (partially) explain the differences found between this Corollary 14 and Theorem 19, consider the expected number of cycles of length 1. This quantity equals $\sum_{i=1}^{N} \frac{a_i b_i}{S} = \frac{\mathbf{a}\cdot\mathbf{b}}{S}$. Since the proof for the upperbound is nearly identical to that of Theorem 19, Corollary 13 and Lemma 11 ,and similar to the proof of our next result, Theorem 21, we omit the details. □

In order to bound $trace(A^r)$ from below with high probability, we will compute the variance of the number of paths of length r that start and end at the same node. We can express $trace(A^r)$, as a summation of indicator random variables for each possible path that could be in a realization of a graph from the Chung-Lu random graph model. Denote each of these indicator variables as X_i. It follows then that

$$Var(trace(A^r)) = Var(\sum X_i) = \sum Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j)$$

. We are primarily interested in pairs i, j such that $Cov(X_i, X_j) \neq 0$ (as we can trivially bound $Var(X_i) = E(X_i^2) - E(X_i)^2 = E(X_i) - E(X_i)^2$ where we used the fact that $E(X_i^2) = E(X_i)$ as X_i is a 0-1 indicator random variable and it follows that $Var(X_i) \leq E(X_i)$).

Theorem 21. *As defined earlier, we denote* $trace(A^r) = \sum_{i=1}^{Z} X_i$ *where* X_i *is an indicator random variable denoting the existence (or lack thereof) of a specific path of length* r *that starts and ends at the same node. Suppose that* $2 < 2r < \frac{\mathbf{a} \cdot \mathbf{b}}{S}$. *It follows that*

$$\sum_{i \neq j} Cov(X_i, X_j) \leq 2r^2 * p_{max} \frac{\mathbf{a} \cdot \mathbf{b}}{S}^{2r-2} exp(\frac{4r^2 p_{max} (\frac{S}{\mathbf{a} \cdot \mathbf{b}})^2}{1 - 2r \frac{S}{\mathbf{a} \cdot \mathbf{b}}}). \tag{4.28}$$

Furthermore it follows that

$$Var(trace(A^r)) \leq E(trace(A^r)) + 2r^2 p_{max} \frac{\mathbf{a} \cdot \mathbf{b}}{S}^{2r-2} exp(\frac{4r^2 p_{max} (\frac{S}{\mathbf{a} \cdot \mathbf{b}})^2}{1 - 2r \frac{S}{\mathbf{a} \cdot \mathbf{b}}}). \tag{4.29}$$

Proof. The proof strategy is analogous to Theorem 19. First consider a path (represented by the indicator variable X_1) of length r that starts and ends at the same node and all of the edges are distinct.

$$Pr(X_1 = 1) = \Pi_{i=1}^{r} p_{x_i x_{i+1}} = \Pi_{i=1}^{r} \frac{b_{x_i} a_{x_{i+1}}}{S} = \Pi_{i=1}^{r} \frac{a_{x_i} b_{x_i}}{S},$$

where the last equality follows from the requirement that $x_1 = x_{r+1}$.

Now consider the case where (only) the mth edge in X_1 is a repeat edge.

It then follows that

$$Pr(X_1 = 1) = \Pi_{i=1, i \neq m}^{r} \frac{b_{x_i} a_{x_{i+1}}}{S} = (\Pi_{i=1}^{m-1} \frac{a_{x_i} b_{x_i}}{S}) \frac{a_{x_m} b_{x_{m+1}}}{S} (\Pi_{i=m+2}^{r} \frac{a_{x_i} b_{x_i}}{S}). \tag{4.30}$$

Or more simply,

$$Pr(X_1 = 1) \leq p_{max} (\Pi_{i=1, i \neq m, m+1}^{r} \frac{a_{x_i} b_{x_i}}{S}). \tag{4.31}$$

Similarly, if $m, m+1, ...m+n$ are all repeat edges it follows that

$$Pr(X_1 = 1) \leq p_{max} (\Pi_{i=1, i \neq [m...m+n+1]}^{r} \frac{a_{x_i} b_{x_i}}{S}). \tag{4.32}$$

And more generally, as in the proof of Theorem 19, we can describe the occurence of repeat blocks as $R = \cup_i [m_i....m_i + n_i + 1]$ where each 'block' or set $[m_i....m_i + n_i + 1]$ is disjoint from all of the other sets in the union.

Furthermore, define B to be the number of 'blocks' or sets in the union that we used to construct R.

It follows in generality that

$$Pr(X_1 = 1) \leq p_{max}^B (\Pi_{i \notin R} \frac{a_{x_i} b_{x_i}}{S}). \tag{4.33}$$

Similarly, we can bound the probability of the existence of two distinct loops,

$$Pr(X_1 = 1, X_2 = 1) \leq p_{max}^B (\Pi_{i \notin R} \frac{a_{x_i} b_{x_i}}{S}) \tag{4.34}$$

where B now represents the total number of repeating node blocks when calculating the probability of the joint existence of paths X_1 and X_2.

Now we are only interested in paths that have a non-trivial covariance. It follows analogous to equation (4.7) in Theorem 19, we have that

$$\sum_{i \neq j} Cov(X_i, X_j) \leq \sum_{\substack{k_1 + \sum_{i=2}^r (i-1) k_i = 2r \\ k_1 \in [1...2r-1]}} \binom{\sum k_i}{k_1, ..., k_r} [k_1^{2r-k_1}] (\frac{\mathbf{a} \cdot \mathbf{b}}{S})^{k_1} \Pi_{i \geq 2} (p_{max} [\frac{S}{\mathbf{a} \cdot \mathbf{b}}])^{k_i}, \tag{4.35}$$

where for $i \geq 2$, the k_i refer to the number of repeating node blocks of length $i - 1$. Using the fact that we can write k_1 in terms of the other k_i, $k_1 = 2r - \sum_{i=2}^r (i-1) k_i$ and that we are only summing over k_i that satisfy the constraint that $k_1 \leq 2r$, we get that

$$\sum_{\substack{k_1 + \sum_{i=2}^r (i-1) k_i = 2r \\ k_1 \in [1...2r-1]}} \frac{(\sum_{i=1}^r k_i)!}{k_1!} (2r)^{2r-k_1} (\frac{\mathbf{a} \cdot \mathbf{b}}{S})^{2r - \sum_{i=2}^r (i-1) k_i} \Pi_{i \geq 2}^r \frac{(\frac{a_{max} b_{max}}{S} [\frac{S}{\mathbf{a} \cdot \mathbf{b}}]^1)^{k_i}}{k_i!}. \tag{4.36}$$

Repeating the same argument from Theorem 19, we achieve the desired upperbound.

$$\sum_{\substack{k_1 + \sum_{i=2}^r (i-1) k_i = 2r \\ k_1 \in [1...2r-1]}} 2r^{\sum_{i=2}^r k_i} (2r)^{\sum_{i=2}^r (i-1) k_i} (\frac{\mathbf{a} \cdot \mathbf{b}}{S})^{2r - \sum_{i=2}^r (i-1) k_i} \Pi_{i \geq 2}^r \frac{(\frac{a_{max} b_{max}}{S} [\frac{S}{\mathbf{a} \cdot \mathbf{b}}]^1)^{k_i}}{k_i!} \leq \tag{4.37}$$

$$(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{2r}\sum_{\substack{k_1+\sum_{i=2}^{r}(i-1)k_i=2r\\ k_1\in[1\ldots 2r-1]}}\Pi_{i\geq 2}^{r}\frac{(\frac{a_{max}b_{max}}{S}[2r\frac{S}{\mathbf{a}\cdot\mathbf{b}}]^i)^{k_i}}{k_i!}. \tag{4.38}$$

Now define $\delta_{ij}=1$ if $i=j$ and zero otherwise. To ease the analysis, we consider all of the different scenarios to identify which k_i ($i\geq 2$) must be at least 1. Once we know which k_i must always be at least 1, since we want to be able to write our upperbound using an exponential function, we will factor out the term so that effectively k_i will be able to range from 0 to $2r$, instead of 1 to $2r$. Since this term is bounded above by $2rp_{max}\frac{S}{\mathbf{a}\cdot\mathbf{b}}$, we get that,

$$2rp_{max}(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{2r-1}\sum_{\substack{k_1+\sum_{i=2}^{r}(i-1)k_i=2r\\ j\in[2,\ldots,r]\\ 1\leq k_j}}\Pi_{i\geq 2}^{r}\frac{(\frac{a_{max}b_{max}}{S}[2r\frac{S}{\mathbf{a}\cdot\mathbf{b}}]^i)^{k_i-\delta_{ij}}}{k_i!}\leq \tag{4.39}$$

$$2rp_{max}(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{2r-1}\sum_{\substack{k_1+\sum_{i=2}^{r}(i-1)k_i=2r\\ j\in[2,\ldots,r]\\ 1\leq k_j}}\Pi_{i\geq 2}^{r}\frac{(\frac{a_{max}b_{max}}{S}[2r\frac{S}{\mathbf{a}\cdot\mathbf{b}}]^i)^{k_i-\delta_{ij}}}{(k_i-\delta_{ij})!}. \tag{4.40}$$

Since there are at most r choices for which k_i is bounded below by 1, we get that,

$$2r^2p_{max}(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{2r-1}\sum_{k_2=0,\ldots,k_r=0}^{\infty}\Pi_{i\geq 2}^{r}\frac{(\frac{a_{max}b_{max}}{S}[2r\frac{S}{\mathbf{a}\cdot\mathbf{b}}]^i)^{k_i}}{k_i!}\leq \tag{4.41}$$

$$2r^2p_{max}(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{2r-1}\Pi_{i\geq 2}^{r}exp(p_{max}[2r\frac{S}{\mathbf{a}\cdot\mathbf{b}}]^i)\leq \tag{4.42}$$

$$2r^2p_{max}(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{2r-1}exp(p_{max}\frac{[2r\frac{S}{\mathbf{a}\cdot\mathbf{b}}]^2}{1-2r\frac{S}{\mathbf{a}\cdot\mathbf{b}}}). \tag{4.43}$$

□

To construct the desired lowerbound on $\rho(A)$, we could appeal to Chebyshev's Inequality. Instead we will use a more distribution specific approach. We will state the result from Janson in full generality and then discuss the implications of their work in context to counting paths and cycles of prescribed length.

Theorem 22. *[Janson 1990] Consider a set of independent random indicator variables* $\{J_i\}_{i\in Q}$ *and a family* $\{Q(\alpha)\}_{\alpha\in B}$ *of subsets of index set* Q. *Define* $I_\alpha = \Pi_{i\in Q(\alpha)} J_i$ *and* $T = \sum_{\alpha\in B} I_\alpha$. *Define* $Cov = \sum_{\alpha_1\neq\alpha_2} Cov(I_{\alpha_1}, I_{\alpha_2})$ *Then for* $\epsilon \in [0,1]$,

$$Pr((1-\beta)E[T] \leq T) \geq 1 - exp(-\frac{1}{2}\frac{(\beta * E[T])^2}{E[T] + Cov}) \tag{4.44}$$

Remark 5. *In the language of Theorem 22 according to the Chung-Lu random graph model, the existence of a particular edge* (J_i) *is independent with respect to the existence of another edge in the graph. As such we define* Q *to be the set that identifies (orders) all of the possible* N^2 *edges that could exist in our graph of* N *nodes. Consider a particular cycle, and denote it by* α. *We can express* α *as a product of independent indicator variables* $\Pi_{i\in Q(\alpha)} J_i$, *where the elements in the set* $Q(\alpha) \subset Q$ *identify the edges (the independent indicator variables* J_i*) used to form the cycle* α. *We denote the indicator variable corresponding to this cycle by* I_α *and let* B *consist of all of the sets* $Q(\alpha)$ *formed by all of the possible cycles* α *of length* r *that could appear in our graph. Hence Theorem 22 provides us with a bound that can make it difficult for the sum of cycles (of prescribed length* r*) to be too much smaller than the expected value. This leads us to the desired result.*

Theorem 23. *Denote* A *as a realization of a random Chung-Lu graph with expected degree sequence* $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{Z}^{N\times 2}$. *Furthermore let* $p_{max} = \frac{a_{max}b_{max}}{S}$ *and* $S = \sum_i a_i = \sum_i b_i$. *Then for every* $\epsilon \in (0,1)$ *there exists a* δ_1,δ_2 *such that if* $\frac{S}{\mathbf{a}\cdot\mathbf{b}} < \delta_1$, $2(logN)^2\frac{S}{\mathbf{a}\cdot\mathbf{b}} < \frac{1}{2}$ *and* $\frac{1}{N} < \delta_2$ *then*

$$Pr((1-\epsilon)\frac{\mathbf{a}\cdot\mathbf{b}}{S} \leq \rho(A)) \geq 1 - \epsilon.$$

Proof. First fix an arbitrary $\epsilon \in (0,1)$ and suppose that

$$(logN)^2\frac{S}{\mathbf{a}\cdot\mathbf{b}} < min(\frac{1}{4}, \sqrt{\frac{-32exp(\frac{1}{2})}{log(\epsilon)}})\ and \tag{4.45}$$

$$log(N) > \frac{6}{\epsilon}. \tag{4.46}$$

Now consider Theorem 22 and consider the number of cycles of length $r = (logN)^2$ which we denote by C. Then we have from Theorem 21 and Lemma 12 that

$$Pr(\frac{E[C]}{2} \leq C) \geq 1 - exp(-\frac{1}{8}\frac{(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{2r}}{[2r^2 p_{max}\frac{\mathbf{a}\cdot\mathbf{b}}{S}^{2r-2} + \frac{\mathbf{a}\cdot\mathbf{b}}{S}^{r}]exp(\frac{4r^2 p_{max}(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2}{1-2r\frac{S}{\mathbf{a}\cdot\mathbf{b}}})}). \tag{4.47}$$

Recall that by assumption (4.45) we have that

$$Pr(\frac{E[C]}{2} \leq C) \geq 1 - exp(-\frac{1}{8}\frac{(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{2}}{[2r^2 p_{max} + \frac{\mathbf{a}\cdot\mathbf{b}}{S}^{-r}]exp(8r^2 p_{max}(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2)}). \tag{4.48}$$

We can simplify this further since $\frac{\mathbf{a}\cdot\mathbf{b}}{S} > 1$ and by assumption (4.45), $(logN)^4 \frac{S}{\mathbf{a}\cdot\mathbf{b}}^2 < \frac{1}{16}$ and say that

$$Pr(\frac{E[C]}{2} \leq C) \geq 1 - exp(-\frac{1}{16}\frac{(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{2}}{[r^2+1]exp(\frac{1}{2})}) \implies \tag{4.49}$$

$$Pr(\frac{E[C]}{2} \leq C) \geq 1 - exp(-\frac{1}{32}\frac{1}{r^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2 exp(\frac{1}{2})}). \tag{4.50}$$

Using assumption (4.45) yields,

$$Pr(\frac{E[C]}{2} \leq C) \geq 1 - \epsilon. \tag{4.51}$$

Finally, with probability $1-\epsilon$ we have that

$$(\frac{1}{2N})^{\frac{1}{(logN)^2}}\frac{\mathbf{a}\cdot\mathbf{b}}{S} \leq (\frac{E(C)}{2N})^{\frac{1}{(logN)^2}} \leq (\frac{C}{N})^{\frac{1}{(logN)^2}} \leq \rho(A), \tag{4.52}$$

where the first inequality comes from Lemma 11 and the final inequality holds from Lemma 10.

Since assumption (4.46) implies that

$$|log(\frac{1}{2N}^{\frac{1}{(logN)^2}})| = |\frac{log(2N)}{(logN)^2}| \leq |\frac{2}{logN}| \leq \frac{\epsilon}{3} \tag{4.53}$$

and $1 - 3x \leq exp(-x) \implies 1 - \epsilon \leq exp(log(\frac{1}{2N})^{\frac{1}{(logN)^2}})$. Consequently from (4.52) with probability at least $1-\epsilon$,

$$(1-\epsilon)\frac{\mathbf{a}\cdot\mathbf{b}}{S} \leq \rho(A).$$

□

Remark 6. *We can in fact get a faster convergence result by considering the number of cycles of length r (where r is a parameter). For $r > 1$ when we invoke Theorem 22, our covariance term C will be positive. For clarity, C_r will denote the cycles of length r and $Cov(r)$ will be used to denote the C mentioned in Theorem 22. We argued in Theorem 21 that $Cov(r) \approx r\frac{S}{\mathbf{a}\cdot\mathbf{b}}^2 * E(C_r)^2$. Since in the limit for r large $E(C_r) \ll E(C_r)^2$ provided $\frac{\mathbf{a}\cdot\mathbf{b}}{S} \to \infty$, we can invoke Theorem 21 where $\beta = \frac{1}{2}$ and show that since $Cov(r) \ll E(C_r)^2$ (for suitably chosen r), with high probability the number of cycles of length r cannot be less than half of $E(C_r)$. Since $\frac{1}{2}\frac{\mathbf{a}\cdot\mathbf{b}}{S}^r \leq C_r \leq \rho(A)^r$, taking the rth root where $r \to \infty$ yields the result. More precisely, if $\frac{\mathbf{a}\cdot\mathbf{b}}{S} = O(N)$, we can prove (a version of) Theorem 23 where choosing $r = O(N)$ implies that $\epsilon = O(\frac{1}{N})$.*

We conclude this section with a simulation (Figure 4.1.2) plotting the empirical distribution of the spectral radius from 100 realizations of the Chung-Lu random graph model from a fixed expected (bi)-degree sequence where there the networks consist of 600 nodes and $\frac{\mathbf{a}\cdot\mathbf{b}}{S} \approx 161$.

4.2 CONCENTRATION BOUNDS WHEN $p_{MAX} \to 0$

We now wish to extend our results to the case where $\mathbf{a}\cdot\mathbf{b}/S$ is (asymptotically) finite. In order to prove results of this nature, we will require that $p_{max} \to 0$, that is the likelihood any two fixed nodes share an edge should vanish asymptotically. We again stress, as suggsested in Figure 4.1.2, that our results not only provide asymptotic information regarding the concentration of the dominating eigenvalue of a sequence of realizations of Chung-Lu random graphs, but also computable concentration results that bound the likelihood that the dominating eigenvalue deviates from $\mathbf{a}\cdot\mathbf{b}/S$ for a randomly generated network from the

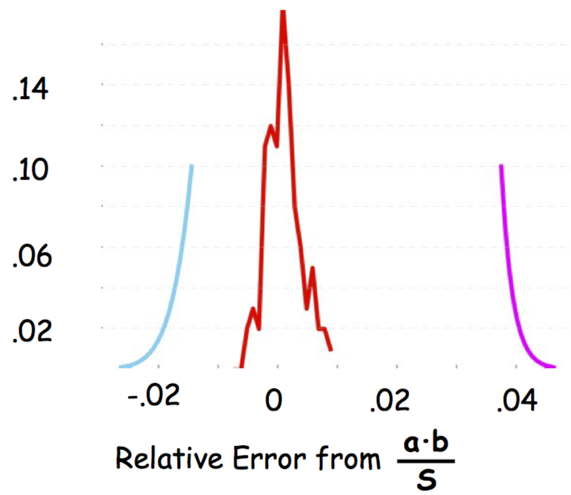

Figure 4.1.2: Speed of the Convergence of $\rho(A)$ to $\frac{\mathbf{a} \cdot \mathbf{b}}{S}$ in the Chung-Lu random graph model. The x-axis indicates the relative error ϵ of the empirically observed $\rho(A)$ such that $\rho(A) = (1+\epsilon)\frac{\mathbf{a} \cdot \mathbf{b}}{S}$. The y axis marks probabilities, whose meanings vary based on the three curves. The red curve is the empirical probability mass function plotting the relative error of the dominating eigenvalue from $\frac{\mathbf{a} \cdot \mathbf{b}}{S}$, where we constructed 100 realizations of Chung-Lu random graphs with a prescribed expected degree sequence $\mathbf{d} \in \mathbb{Z}^{600 \times 2}$ such that $\frac{\mathbf{a} \cdot \mathbf{b}}{S} \approx 161$. The magenta curve is an application of the concentration result (Theorem 20) regarding the distribution of the dominating eigenvalue $\rho(A)$ from realizations of Chung-Lu random graphs with the same prescribed expected degree sequence $\mathbf{d}$. For this curve, the $y - axis$ provides an upperbound on the probability the spectral radius exceeds the relative error from $\frac{\mathbf{a} \cdot \mathbf{b}}{S}$ on the x-axis. Similarly, the blue curve is an application of the concentration result from Theorem 23.

Chung-Lu random graph model with a fixed number of nodes. The following definition will help us greatly in achieving our desired goal.

Definition 8. *A simple cycle is a path that begins and ends at the same node where no other node is visited more than once. Denote the number of simple cycles of length r by $SC(r)$.*

Since the number of simple cycles of length r is a lower bound for the $trace(A^r)$, it suffices to show that with high probability that the number of simple cycles of length r is roughly $(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^r$.

Lemma 12. *Consider a realization of the Directed Chung-Lu random graph model with expected degree sequence $\mathbf{d} = (\mathbf{a}, \mathbf{b})$, where $\sum a_i = \sum b_i = S$. Then*

$$[\frac{\mathbf{a}\cdot\mathbf{b}}{S} - rp_{max}]^r \leq E(SC(r)) \leq (\frac{\mathbf{a}\cdot\mathbf{b}}{S})^r. \tag{4.54}$$

Proof. We can write the expected number of simple cycles of length r as,

$$E(SC(r)) = \sum_{i_1 = i_{r+1} \neq i_2 \ldots \neq i_r} \Pi_{k=1}^r p_{i_k i_{k+1}} = \sum_{i_1 \neq \ldots \neq i_r} \Pi_{k=1}^r \frac{a_{i_k} b_{i_k}}{S}.$$

where the first equality follows from the fact that with simple cycles we do not have to worry about an edge repeating in a cycle (as in Lemma 11) and the second equality follows from invoking the definition of Chung-Lu and rearranging terms.

We then have the upperbound that,

$$E(SC(r)) \leq \sum_{i_1,\ldots,i_r} \Pi_{k=1}^r \frac{a_{i_k} b_{i_k}}{S} = (\sum_{i_1} \frac{a_{i_1} b_{i_1}}{S})(\sum_{i_2,\ldots,i_r} \Pi_{k=2}^r \frac{a_{i_k} b_{i_k}}{S}) = (\frac{\mathbf{a}\cdot\mathbf{b}}{S})(\sum_{i_2,\ldots,i_r} \Pi_{k=2}^r \frac{a_{i_k} b_{i_k}}{S}) = (\frac{\mathbf{a}\cdot\mathbf{b}}{S})^r.$$

To derive the lower bound,

$$(\frac{\mathbf{a}\cdot\mathbf{b}}{S} - rp_{max}) \sum_{i_1 \neq \ldots \neq i_{r-1}} \Pi_{k=1}^{r-1} \frac{a_{i_k} b_{i_k}}{S} = (\frac{\mathbf{a}\cdot\mathbf{b}}{S} - r\frac{a_{max} b_{max}}{S}) \sum_{i_1 \neq \ldots \neq i_{r-1}} \Pi_{k=1}^{r-1} \frac{a_{i_k} b_{i_k}}{S} \leq \sum_{i_1 \neq \ldots \neq i_r} \Pi_{k=1}^r \frac{a_{i_k} b_{i_k}}{S}.$$

Repeating this argument for each of the i_k's yields that,

$$(\frac{\mathbf{a}\cdot\mathbf{b}}{S} - rp_{max})^r \leq \sum_{i_1 \neq \ldots \neq i_r} \Pi_{k=1}^r \frac{a_{i_k} b_{i_k}}{S}.$$

□

We will also want to construct our covariance term as used in Theorem 22. To do so, we have the following result.

Lemma 13. *Consider the covariance term for the number of simple cycles of length r, denoted by $Cov(r)$, (as defined in Theorem 22 where our covariance term is derived from the expectation of sum of products consisting of pairs of indicator variables, the existence of simple cycles).*

We then have that

$$Cov(r) \leq p_{max} r^2 (\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{2r-2} exp(p_{max} r^2 \frac{\frac{S}{\mathbf{a}\cdot\mathbf{b}}}{1-\frac{S}{\mathbf{a}\cdot\mathbf{b}}}). \tag{4.55}$$

Proof. In the spirit of Theorems 19 and 21, we claim that the following is an upperbound for the covariance term.

$$Cov(r) \leq \sum_{\substack{k_1+\sum_{i=2}^{r}(i-1)k_i=2r \\ 1\leq k_1\leq 2r-1}} \binom{\sum k_i}{k_1,...,k_r} (\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{k_1} \Pi_{i\geq 2} (p_{max}[r\frac{S}{\mathbf{a}\cdot\mathbf{b}}]^1)^{k_i} \tag{4.56}$$

The key difference between (4.56) and the claim in Theorem 21, is that because we are only considering simple cycles, when evaluating the covariance of two distinct simple cycles each with length r, in the first cycle there cannot be a repeated node block (because it is a simple cycle) and in the second cycle if there is a repeated node block of size i, then you have at most r choices for constructing that repeated node block (as opposed to r^i).

To clarify this point, in the second simple cycle suppose a repeated node block of size z appears of the form $m_1 \to m_2 \to ... \to m_z$. (Consider a particular choice for m_1.) Since m_1 can only appear once in a simple cycle, it must be that there is only one possible choice for m_2, the node that appears after m_1 in the first cycle. Proceeding inductively, given a particular value for m_1, all of the other $m's$ are completely determined. Hence there are at most r distinct choices for each block of size z. Employing the same tricks as in Theorem 19, demonstrates that,

$$\sum_{\substack{k_1+\sum_{i=2}^{r}(i-1)k_i=2r \\ 1\leq k_1\leq 2r-1}} (\sum_{i=1}^{r} k_i)^{\sum_{i=2}^{r} k_i} (\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{2r-1-\sum_{i=2}^{r}(i-1)k_i} \Pi_{i\geq 2} \frac{(p_{max}[r\frac{S}{\mathbf{a}\cdot\mathbf{b}}])^{k_i}}{k_i!} \leq \tag{4.57}$$

$$(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{2r-1}\sum_{\substack{k_1+\sum_{i=2}^{r}(i-1)k_i=2r\\ 1\le k_1\le 2r-1}}(r)^{\sum_{i=2}^{r}k_i}\Pi_{i\ge 2}\frac{(p_{max}[r(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^i])^{k_i}}{k_i!}\le \tag{4.58}$$

$$(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{2r-1}\sum_{\substack{k_1+\sum_{i=2}^{r}(i-1)k_i=2r\\ 1\le k_1\le 2r-1}}\Pi_{i\ge 2}\frac{(p_{max}[r^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^i])^{k_i}}{k_i!}. \tag{4.59}$$

Now since k_1 cannot exceed $2r-1$, at least one of the $k_i's$ $(i \geq 2)$ is positive. Hence inside the summation, all of these terms have a common factor of $p_{max}r^2\frac{S}{\mathbf{a}\cdot\mathbf{b}}$. It then follows,

$$p_{max}r^2(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{2r-2}\sum_{k_2=0,\ldots,k_r=0}^{\infty}\Pi_{i\ge 2}\frac{(p_{max}[r^2(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^i])^{k_i}}{k_i!}\le \tag{4.60}$$

$$p_{max}r^2(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{2r-2}exp(p_{max}r^2\frac{\frac{S}{\mathbf{a}\cdot\mathbf{b}}}{1-\frac{S}{\mathbf{a}\cdot\mathbf{b}}}). \tag{4.61}$$

And since (4.61) is an upperbound for $Cov(r)$, the proof is complete. □

Consequently, we have the following concentration result.

Corollary 15. *Denote A as a realization of a random Chung-Lu graph with expected degree sequence $\mathbf{d} = (\mathbf{a}, \mathbf{b}) \in \mathbb{Z}^{N\times 2}$. Furthermore let $p_{max} = \frac{a_{max}b_{max}}{S}$ and $S = \sum_i a_i = \sum_i b_i$. Then for every $\epsilon \in (0, 1)$ there exists a δ_1,δ_2 such that if $p_{max}(logN)^2 < \delta_1$ and $\frac{1}{N} < \delta_2$ then*

$$Pr((1-\epsilon)\frac{\mathbf{a}\cdot\mathbf{b}}{S}\le\rho(A))\ge 1-\epsilon.$$

Proof. The proof is analogous to Theorem 23 and as such we only provide a sketch. From Lemmas 13 and 12, for sufficiently small p_{max}, we have that $Cov(SC((logN)^2) \ll [E(SC((logN)^2)]^2$. As such we can invoke Theorem 22 to show that with high probability the number of simple cycles of length $(logN)^2$ cannot be less than half the expected number of simple cycles.

$$\frac{E(SC((logN)^2))}{2N}\le\frac{SC((logN)^2)}{N}\le\frac{trace(A^{(logN)^2})}{N}\le\rho(A)^{(logN)^2}.$$

As $\frac{(\frac{\mathbf{a}\cdot\mathbf{b}}{S}-(logN)^2p_{max})^{(logN)^2}}{2N}\le\frac{E(SC((logN)^2))}{2N}$, this would impliy that,

$$(\frac{1}{2N})^{\frac{1}{log(N)^2}}(\frac{\mathbf{a}\cdot\mathbf{b}}{S}-log(N)^2p_{max})\le\rho(A).$$

Since $(\frac{1}{2N})^{\frac{1}{(logN)^2}} \to 1$ for N sufficiently large and $(logN)^2 p_{max} \to 0$, we conclude that with high probability $(1-\epsilon)$,

$$(1-\epsilon)\frac{\mathbf{a}\cdot\mathbf{b}}{S} \leq \rho(A).$$

□

Constructing a meaningful upperbound on the number of paths of length r when $p_{max} \to 0$ is more challenging. We cannot simply follow the proof strategy used in Theorem 19 as we have to worry about the contribution from repeating node blocks. Analogously, consider the (unlikely) event that a subgraph of k nodes exists where each node has bidirectional edges with each of the other k nodes and $\frac{\mathbf{a}\cdot\mathbf{b}}{S} \ll k$. In the proof of Theorem 19, we were able to ignore this complication altogether due to the assumption that $\frac{\mathbf{a}\cdot\mathbf{b}}{S} \to \infty$ sufficiently fast. The following lemma demonstrates that with high probability when $p_{max} \to 0$,we in fact do not have to worry (too much) about this dilemna.

Definition 9. *Define the minimal edge set of a path P to be an ordered list of edges (of minimal size) required for the entire path to exist. Note that since P must have minimal size, we cannot have the same edge appear twice in the minimal edge set. By convention when constructing a minimal edge set, as we observe edges in a path, we simply add the edge to the minimal edge set if we have never observed that particular edge before. Furthermore, we say that a path has k node repetitions if there are k distinct simple cycles that can be formed from the minimal edge set. Analogously, we can say that an edge set has k node repetitions if there are k distinct simple cycles that can be formed from that edge set.*

Example: Consider the path $P = 1 \to 2 \to 3 \to 1 \to 2 \to 4$. For the path P to exist, we only need the following edges to exist: $1 \to 2$, $2 \to 3$, $3 \to 1$ and $2 \to 4$. These edges $\{(1,2),(2,3),(3,1),(2,4)\}$ would then form the minimal edge set of path P.

Essentially, we will seek a result that says that paths cannot have too many node repetitions. More precisely, it will be burdensome to consider the minimal edge set of a path P with excessively many node repetitions. Instead, we will want to construct a subset (of edges) of the minimal edge set with a smaller number of node repetitons. The following Lemma formalizes this claim and the proof explains how to construct such a subset.

Lemma 14. *Consider a path P with $k > 1$ node repetitions and its corresponding minimal edge set M. Then for any $j \leq k$, we can construct a* ***reduced edge set*** *$E \subset M$, such that E has j node repetitions, the first node in the first edge in E and the last node in the last edge in E must belong to a simple cycle that can be formed from the edges in E.*

Proof. We will illustrate the procedure using an example, but the procedure used holds in the full generality of the lemma statement.

Consider the path $P = 1 \to 2 \to 2 \to 3 \to 4 \to 3 \to 1$ and its minimal edge set $M = \{(1,2), (2,2), (2,3), (3,4), (4,3), (3,1)\}$.

- First, starting with the empty set we add edges from our minimal edge set M to our reduced edge set E until we can construct j simple cycles. For this example, consider $j = 2$. This yields the edge set $E = \{(1,2), (2,2), (2,3), (3,4), (4,3)\}$.
- Then, we remove edges from the beginning of our edge set E until we reach an edge such that its removal would decrease number of simple cycles that we can form our current edge set. This yields the set $E = \{(2,2), (2,3), (3,4), (4,3)\}$.
- Observe that if any of the edges in the edge set E do not exist, then the path P cannot exist. Furthermore by construction, the first and last node in our new edge set must belong to a simple cycle that we can construct from our edge set.

□

With Lemma 14 at hand, we seek one more lemma so that we can prove the desired result that there cannot exist a path with many node repetitions with high probability.

Lemma 15. *Suppose E is a reduced edge set with t node repetitions, as we constructed in Lemma 14. We can map the reduced edge set E to a union of sets $\cup_{i=1}^{t} M_i$ where each M_i consists of a set of nodes. Furthermore, the M_i sets have the following properties:*

- *The first node in M_1 and the last node in M_t all belong to a simple cycle that can be constructed from the edges in E (provided that the edges in E exist).*

- *The last node in M_j for each j is a node that has appeared either before in the same set M_j or in a set M_i where $i < j$.*
- *The first node in M_j for each $j > 1$ is a node that has appeared in some set M_i where $i < j$.*
- *And finally, let $x_{k,i}$ be the kth node in M_i. Then the probability that all edges in the reduced edge set E exist equals*

$$\Pi_{i=1}^{t}\Pi_{k=1}^{|M_i|-1}\frac{b_{x_{k,i}}a_{x_{k+1,i}}}{S}. \tag{4.62}$$

Proof. We first illustrate the mapping of the reduced edge set E with t node repetitions to the union of sets $\cup_{i=1}^{t}M_i$ where each M_i consists of a set of nodes. We start by decomposing our reduced edge set E as a union of edge sets $\cup_{i=1}^{t}E_i$, where we add the edges from E to E_1 and stop once E_1 has precisely one node repetition. Then starting where we left off, we add edges to E_2 and stop once $E_1 \cup E_2$ have precisely two node repetitions. Since by assumption the last node of the last edge belongs to a simple cycle, we can express $E = \cup_{i=1}^{t}E_i$ such that the subset $\cup_{i=1}^{k}E_i$ has precisely k node repetitions for all $k \leq t$.

Example: We clarify the above procedure with the following example. Consider the reduced edge set $E = \{(2,2),(2,3),(3,4),(4,3)\}$. Then since E has two node repetitions and the first node of the first edge and the last node of the last edge both belong to simple cycles that can be formed by the edges of E, we can write $E_1 = \{(2,2)\}$, since E_1 already has one node repetition we stop here and then $E_2 = \{(2,3),(3,4),(4,3)\}$. Note that E_1 has precisely one node repetition and $E_1 \cup E_2 = E$ has two node repetitions.

Now given E_i, we can construct M_i as follows. Consider the first node of each edge in E_i and add those nodes to M_i in that order. Then add the last node of the last edge of E_i to M_i. This completes the construction of the M_i.

Example: As before consider $E = E_1 \cup E_2 = \{(2,2)\} \cup \{(2,3),(3,4),(4,3)\}$. It then follows that $M_1 = \{2,2\}$, where we added the first node of each edge of E_1 and then the last node of the last edge of E_1 to M_1 and $M_2 = \{2,3,4,3\}$. Now note that the probability of all of the edges in E existing equals $\Pi_{i=1}^{2}\Pi_{k=1}^{|M_i|-1}\frac{b_{x_{k,i}}a_{x_{k+1,i}}}{S}$.

Note that by construction, since the first node of the first edge and the last node of the last edge in E belong to a simple cycle that can be constructed from the edges of E, these nodes are precisely the first and last nodes of M_1 and M_t respectively. Hence the first bulleted statement holds.

The second bulleted statement holds since by construction $\cup_{i=1}^{k}E_i$ the last node in the last edge of E_k forms a simple cycle. Consequently, this node appears as the last node in M_k and must have appeared elsewhere in the set $\cup_{i=1}^{k}M_i$.

Now for the third bulleted statement, we have two cases; either the first node in M_i equals the last node in M_{i-1}, in which the claim is trivial, or they are distinct. If they are distinct, since each minimal edge set is constructed from a path, then in the path there exists a collection of edges in the minimal edge set M that connect the last node in M_{i-1} to the first node in M_i. Denote these nodes as m_{i-1} and m_i respectively.

It follows from the second bulleted statement that there is a simple cycle containing m_{i-1}. Now consider the collection of edges that appear earlier in M that connect m_{i-1} to m_i. Note that if we delete an edge that belongs to a simple cycle in a minimal edge set, the number of node repetitions decreases by 1, as all edges in the minimal edge set can only appear once. Recall that in the proof of Lemma 14, to form the reduced edge set E, if $M = \{e_1, ..., e_r\}$, then $E = \{e_j, e_{j+1}, ..., e_k\}$ for some j, k such that $1 \leq j \leq k \leq r$. Now we only delete edges in the beginning, $e_1, ..., e_{j-1}$ if those edges do not contribute to the number of node repetitions. Since by construction the appearance of the edges that connect m_{i-1} to m_i must appear before e_k and the earlier appearance of the edges connecting m_{i-1} to m_i would contribute to the number of node repetitions, we conclude that the path that connects m_{i-1} to m_i does appear earlier in E and consequently m_i appears earlier in some M_{i_0} for $i_0 < i$.

Denote the probability that all edges in E exist by $Pr(E)$. It follows that since E is a subset of the minimal edge set of a path P all of the edges are distinct and that

$$Pr(E) = \Pi_{e=(x,y)\in E}\frac{b_x a_y}{S} = \Pi_{i=1}^{t}\Pi_{e=(x,y)\in E_i}\frac{b_x a_y}{S}. \quad (4.63)$$

Now consider the kth edge in E_j for some particular choice of j. Denote this edge as (x, y). It follows from construction of M_j that x is precisely the kth element in M_j and y is the $k+1st$ element in M_j. Consequently, we conclude that

$$Pr(E) = \Pi_{i=1}^{t}\Pi_{e=(x,y)\in E_i}\frac{b_x a_y}{S} = \Pi_{i=1}^{t}\Pi_{k=1}^{|M_i|-1}\frac{b_{x_{k,i}}a_{x_{k+1,i}}}{S}, \quad (4.64)$$

where $x_{k,i}$ is the kth node in M_i.

□

The strength of Lemma 15 lies in the fact that it helps us identify which nodes repeat in a reduced edge set. For example consider some arbitrary bounded function $f : \mathbb{N} \to \mathbb{R}$ and define f_* to be the smallest upperbound of f. Then it follows that, $\sum_{i=1,j=1}^{N}[f(i)]^2 f(j) \leq f_* \sum_{i=1,j=1}^{N} f(i)f(j)$; if $f_* \to 0$ and $\sum_{i=1}^{N} f(i) = O(1)$, then this upperbound converges to 0. To summarize, we will want to identify which indices (nodes) repeat in the reduced edge set as if for example we chose the wrong index and bounded $\sum_{i=1,j=1}^{N}[f(i)]^2 f(j)$ by $f_* \sum_{i=1,j=1}^{N}[f(i)]^2 \leq N f_* \sum_{i=1}^{N}[f(i)]^2$, the presence of the N would lead us to a potentially useless upperbound as it is possible that $f_* \to 0$ and $\frac{1}{f_*^2} \ll N$. By noting that with a little care we can treat this abstract function f in place of the probabilities used to construct the edges in our graph, we can derive the following result.

Lemma 16. *Consider a sequence of (expected) degree sequences* $\mathbf{d} = (\mathbf{a}, \mathbf{b})$ *where* $p_{max} \leq \frac{R}{N^{\tau}}$*, R is a fixed constant, $\tau > 0$ and $\frac{\mathbf{a}\cdot\mathbf{b}}{S} > 1$. Then with probability at least $p_* = 1 - \delta$, all paths of length not exceeding $L = \frac{k\tau}{2} * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)$, have less than $k+1$ node repetitions, where* $\delta = \frac{R^k L^{3k-2}}{N^{\frac{k\tau}{2}}}$.

Remark: Note that asymptotically we are guaranteed that $\delta \to 0$ in Lemma 16 as for fixed k, $p_{max} = O(N^{-\tau})$, $L = O(log(N))$ and consequently $\delta = O((N^{\frac{-\tau}{2}}(logN)^3)^k) \to 0$.

Proof. To bound the likelihood of the existence of any path of length at most L with at least t node repetitions, we will instead consider the likelihood of the existence of a reduced edge set with t node repetitions containing no more than $L+1-t$ distinct nodes. Denote $Pr(E_*)$ as the probability all edges in the set E_* exist. It follows from Lemma 15 that for a given reduced edge set E_* with t node repetitions that

$$Pr(E_*) = \Pi_{i=1}^{t}\Pi_{k=1}^{|M_i|-1}\frac{b_{x_{k,i}}a_{x_{k+1,i}}}{S} = \Pi_{i=1}^{t}\frac{b_{x_{1,i}}a_{x_{|M_i|,i}}}{S}\Pi_{k=2}^{|M_i|-1}\frac{b_{x_{k,i}}a_{x_{k,i}}}{S} = \tag{4.65}$$

$$\frac{b_{x_{1,1}}a_{x_{|M_t|,t}}}{S}\Pi_{j=1}^{t-1}\frac{a_{x_{|M_j|,j}}b_{x_{1,j+1}}}{S}\Pi_{i=1}^{t}\Pi_{k=2}^{|M_i|-1}\frac{b_{x_{k,i}}a_{x_{k,i}}}{S} \tag{4.66}$$

where $x_{k,i}$ is the kth node in M_i.

Ideally, we would like to bound each of the $\frac{a_{x_{|M_j|,j}}b_{x_{1,j+1}}}{S}$ terms by a p_{max}. However, we only want to do so when $x_{1,j}$ and $x_{|M_j|,j}$ already contribute a term of the form $\frac{b_{x_*}a_{x_*}}{S}$ so that when we sum over all possible choices for the nodes in $\cup_{i=1}^{t}M_i$, we do not end up with an extra factor of an N, where N is the number of nodes. Normally, this is not a problem as we know from Lemma 15 that the last node for every M_j can be found either in the same M_j or an earlier M_i where $i < j$. Similarly, the first node in every M_j, except for M_1 can also be found in an earlier M_i where $i < j$. Consequently, we focus on whether the first node in M_1 appears elsewhere.

To analyze (4.65) we have the following cases.

- Case 1: The first node in M_1 and the last node in M_t are different.
  - 1a. Since the first node and last node must respectively be the first and last node in two different simple cycles by Lemma 14, the first and last node must appear elsewhere in the $M_k's$. Case 1a. supposes that there exists an M_{k-1}, M_k such that the last node in M_{k-1} equals the first node in M_k which equals the first node in M_1.
  - 1b. Alternatively, there exists an M_{k-1} such that the last node in M_{k-1} equals the first node in M_1, but the first node in M_k does not equal the first node in M_1.
- Case 2: The first node in M_1 equals the last node in M_t.

First in Case 1a, we know that there exists a k, such that for the first node $x_{1,1} = x_{|M_k|,k} = x_{1,k+1}$. Similarly for the last node we know there exists a k_0 and k_1 such that $1 < k_0 < |M_{k_1}|$

and $(k_0, k_1) \neq (|M_t|, t)$ where $x_{k_0,k_1} = x_{|M_t|,t}$ as by our representation of our reduced edge set as a union of sets of $M_k's$, the first time a node appears (except for the very first node in our reduced edge set), it cannot appear at the beginning of an M_k. Consequently, since Lemma 14 requires that $x_{|M_t|,t}$ is part of a simple cycle and by assumption $x_{1,1} \neq x_{|M_t|,t}$, there exists a distinct (k_0, k_1) such that $x_{k_0,k_1} = x_{|M_t|,t}$. Consequently in Case 1a, we can bound $b_{x_{1,1}}$ by b_{max} and $a_{x_{|M_t|,t}}$ by a_{max} and we are guaranteed that a factor of $\frac{b_{x_*}a_{x_*}}{S}$ appears in (4.65) for $x_* = x_{1,1}$ and $x_* = x_{|M_t|,t}$. For Case 1a we have the following upperbound,

$$\frac{b_{x_{1,1}}a_{x_{|M_t|,t}}}{S}\frac{b_{x_{1,1}}a_{x_{1,1}}}{S}\Pi_{j=1,j\neq k}^{t-1}\frac{a_{x_{|M_j|,j}}b_{x_{1,j+1}}}{S}\Pi_{i=1}^{t}\Pi_{k=2}^{|M_i|-1}\frac{b_{x_{k,i}}a_{x_{k,i}}}{S} \leq p_{max}^{t-1}\frac{b_{x_{1,1}}a_{x_{1,1}}}{S}\Pi_{i=1}^{t}\Pi_{k=2}^{|M_i|-1}\frac{b_{x_{k,i}}a_{x_{k,i}}}{S}. \tag{4.67}$$

where everytime we apply the p_{max} upperbound to the product of degrees of nodes, those nodes also appear in the product $\frac{b_{x_{1,1}}a_{x_{1,1}}}{S}\Pi_{i=1}^{t}\Pi_{k=2}^{|M_i|-1}\frac{b_{x_{k,i}}a_{x_{k,i}}}{S}$ in (4.67). Note that to derive (4.67) $\Pi_{i=1,i\neq k}^{t-1}\frac{a_{x_{|M_i|,i}}b_{x_{1,i+1}}}{S}$ contributes a p_{max}^{t-2} and $\frac{b_{x_{1,1}}a_{x_{|M_t|,t}}}{S}$ contributes a factor of p_{max}.

Now in Case 1b, we again have that for the last node we know there exists a k_1 and k_2 such that $1 < k_2 < |M_{k_1}|$ such that $x_{k_2,k_1} = x_{|M_t|,t}$. Furthermore by assumption, we know that there exists a $k \neq t$ such that $x_{1,1} = x_{|M_k|,k}$ and $x_{k+1,1} \neq x_{1,1}$. We conclude that $x_{k+1,1}$ must have appeared elsewhere and it follows from (4.65) that,

$$\frac{b_{x_{k+1,1}}a_{x_{|M_t|,t}}}{S}\frac{b_{x_{1,1}}a_{x_{|M_k|,k}}}{S}\Pi_{j=1,j\neq k}^{t-1}\frac{a_{x_{|M_j|,j}}b_{x_{1,j+1}}}{S}\Pi_{i=1}^{t}\Pi_{k=2}^{|M_i|-1}\frac{b_{x_{k,i}}a_{x_{k,i}}}{S} \leq \tag{4.68}$$

$$p_{max}^{t-1}\frac{b_{x_{1,1}}a_{x_{1,1}}}{S}\Pi_{i=1}^{t}\Pi_{k=2}^{|M_i|-1}\frac{b_{x_{k,i}}a_{x_{k,i}}}{S}.$$

Finally in Case 2, we know the first node equals the last node $x_{1,1} = x_{|M_t|,t}$. We then can reduce (4.65) to

$$\frac{b_{x_{1,1}}a_{x_{1,1}}}{S}\Pi_{j=1}^{t-1}\frac{a_{x_{|M_j|,j}}b_{x_{1,j+1}}}{S}\Pi_{i=1}^{t}\Pi_{k=2}^{|M_i|-1}\frac{b_{x_{k,i}}a_{x_{k,i}}}{S} \leq \frac{b_{x_{1,1}}a_{x_{1,1}}}{S}p_{max}^{t-1}\Pi_{i=1}^{t}\Pi_{k=2}^{|M_i|-1}\frac{b_{x_{k,i}}a_{x_{k,i}}}{S} \tag{4.69}$$

Denote the set $\mathbf{E}_t$ to contain all of the reduced edge sets E_* with t node repetitions and no more than $L+1-t$ distinct nodes, where $t \geq 2$. We implicitly denote the dependence of

M_i on E_* and define $Pr(\mathbf{E}_t)$ to be the probability that there exists an $E_* \in \mathbf{E}_t$ such that all of the edges in E_* exist..

Then from (4.65),(4.67),(4.68) and (4.69) we conclude that

$$Pr(\mathbf{E}_t) \leq \sum_{E_* \in \mathbf{E}_t} \frac{b_{x_{1,1}} a_{x_{1,1}}}{S} p_{max}^{t-1} \Pi_{i=1}^{t} \Pi_{k=2}^{|M_i|-1} \frac{b_{x_{k,i}} a_{x_{k,i}}}{S} \leq \tag{4.70}$$

$$L^{3t-2} p_{max}^{t-1} (\frac{\mathbf{a} \cdot \mathbf{b}}{S})^{\sum |M_i| - 2t} \leq L^{3t-2} p_{max}^{t-1} (\frac{\mathbf{a} \cdot \mathbf{b}}{S})^{L}. \tag{4.71}$$

where we get $(\frac{\mathbf{a} \cdot \mathbf{b}}{S})$ by summing over all possible choices for the $x_{i,j}$. Furthermore we get a factor of L^{2t-2} as we can choose at most L different nodes to be $x_{k,1}$ for each $k > 1$ and $x_{k,|M_k|}$ and each $k < t$. We then bound the different possible choices for the sizes of the t M_k sets by L^t. Choose $L = \frac{k\tau}{2} log_{\frac{\mathbf{a} \cdot \mathbf{b}}{S}}(N)$. Note that $(\frac{\mathbf{a} \cdot \mathbf{b}}{S})^L = N^{\frac{k\tau}{2}}$.

Hence it then follows from (4.71) that

$$Pr(\mathbf{E}_t) \leq \frac{L^{3t-2} N^{\frac{t\tau}{2}} R^t}{N^{t\tau}} = \frac{L^{3t-2} R^t}{N^{t\tau}}. \tag{4.72}$$

where we invoked the fact that $p_{max} \leq \frac{R}{N^\tau}$. Note that if we choose $L = O((logN)^2)$, since R is a constant, $Pr(\mathbf{E}_t) \to 0$ and hence if we fix t, the likelihood of the existence of a path of length no greater than L with t node repetitions vanishes asymptotically as $N \to \infty$. □

Lemma 16 asymptotically guarantees that with high probability any path in a realization of a Chung-Lu graph of less than $O(log(N))$ length cannot contain more than one distinct simple cycle. Such a restriction will help us in counting the number of repeating node blocks. Suppose we are guaranteed that there are a total of m node repetitions in a given path. If we look at all potential repeating node blocks of less than $O(log(N))$ length, we know that we can only place one distinct simple cycle in the entire node block. At this juncture, we consider Figure 4.2.1 to help us identify the paths that Lemma 16 says cannot exist with high probability.

In each section of Figure 4.2.1, we used colored rectangles to indicate the presence of a node on a given path. For clarity the numbers within each of the rectangles indicate the node's identity. The unfilled spaces before, after and in between the colored rectangles

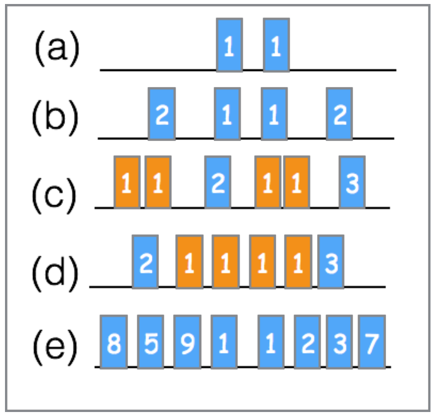

Figure 4.2.1: Restriction on Repeating Edge Blocks for $p_{max} \rightarrow 0$. Each subcase (a-e) indicates a collection of paths of sufficiently small length that may or may not exist in a Chung-Lu random graph with high probability according to Lemma 16. The numbers within each of the rectangles indicate the node's identity. We denote certain nodes using orange rectangles, as opposed to blue rectangles, to help identify when a simple cycle within a particular path beings and ends. In case (a), we are considering a collection of paths that contain a cycle formed by the node labeled 1. Subsequently in case (b), we consider a single path that contains two cycles, one of the cycles begins and ends with node 1 and another that contains node 2. In case (c), we consider two distinct cycles that contain the node 1. Note that the color orange helps distinguish between the two different cycles. The first cycle is relatively small and does not contain the node 2, while the second cycle does contain node 2. Case (d) illustrates the possibility that a single path could contain a concatenation of the same cycle. And finally in case (e), we illustrate a subset of the paths considered in case (a), where we consider specific nodes to the left and right of the cycle containing node 1. See the text for details.

(nodes) exemplify the idea that even without knowing all of the nodes in the path, we can still invoke Lemma 16 to rule out the likelihood of its existence. The key observation is that wherever we place the given distinct cycle, we cannot have additional different cycles in a node block/path of sufficiently small length.

In part (a) of Figure 4.2.1 we decided the location of a particular cycle by visiting node 1 twice in two particular locations. But if we carelessly assign locations for additional nodes and create another cycle, as indicated in Part (b) of Figure 4.2.1, this violates Lemma 16 because now we have two node repetitions (distinct cycles) within a path bounded by $O(log(N))$ length.

Similarly (c) is another violation of Lemma 16. Here, as identifying when a cycle begins and ends can be challenging, we used the color orange to emphasize the start and completion of a cycle. We have two distinct cycles, one cycle from the first two orange $1's$ and a longer second cycle where the blue 2 is in between the two orange $1's$.

However, as demonstrated in part (d), we can concatenate the same cycle over and over again. This will not be a violation of Lemma 16. Ultimately though once we decide the location of our cycle and the number of times we concatenate the same cycle to itself, we cannot have another node repetition as illustrated in part (e) of Figure 4.2.1.

We now provide some intuition for how Lemma 16 helps us limit the number of repeating node blocks within a given path. First consider a path of fixed length r and break up the path into portions each of length $O(log(N))$. Now whenever a repeating node block occurs we have a limited number of choices on how to choose which nodes appear in the repeating node blocks. Recall how we mapped our reduced edge set to a union of sets of nodes, $\cup M_i$ in Lemma 15. Note that we are constructing our repeating node block based on edges already visited, so consider the auxilliary graph induced by these edges. Only nodes that appear at the end or the beginning of a set M_i can have more than one incoming or outgoing neighbor in this auxiliary graph. More precisely, if the path has at most m node repetitions, we can envision the construction of our auxiliary graph such that each node has 1 inward and outgoing edge and then m additional edges are added based on the nodes that appear at the end and beginning of the M_i sets. Now from Lemma 16 we know that there cannot be more than two distinct simple cycles in an $O(log(N))$ block of any given path. Hence after we decide the location of a particular cycle in this block, one can show that if a path has at most m node repetitions, then before the concatenation of the particular cycles there are at most $m^k exp(m)$ possible ways of choosing nodes for k repeating nodes blocks of arbitrary size that are contained within this $O(log(N))$ portion of the path. We achieve this (non-trivial) bound as there are m^k choices for deciding the inital choice in each repeating node block (based on the constrained number of node repetitions in the path), When choosing nodes in the repeating node blocks, we are restricted by the edges found in the auxiliary graph, where most nodes have only one neighbor. As such one can show that, we need to multiply the quantity m^k by $exp(m)$ as Lemma 16 prevents us from having many different nodes that have more than degree 1 in the auxiliary graph. With this idea at hand, we now present the desired Theorem.

Theorem 24. *Consider a sequence of (expected) degree sequences* $\mathbf{d} = (\mathbf{a}, \mathbf{b})$ *where* $p_{max} \leq \frac{R}{N^{\tau}}$*, R is a fixed constant, $\tau > 0$ and* $\frac{\mathbf{a}\cdot\mathbf{b}}{S} > 1$*. Let* $r = \frac{m\tau}{2} * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)$*. Define* $P_r(y, \mathbf{G})$ *to be the number of paths of length r starting at node y for a given graph* $G \in \mathbf{G}$*, where no path of length r in* $\mathbf{G}$ *has more than $m+1$ node repetitions and no path in* $\mathbf{G}$ *of length less than* $\frac{\tau}{2} * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)$ *has more than 1 node repetition. (Recall Lemma 16). For notational simplicity define*

$$\eta = [\tau \; log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)]^2(m+1)^2$$

and suppose that

$$(\eta)^{\frac{2}{\tau * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)}} < \frac{\mathbf{a}\cdot\mathbf{b}}{S},$$

then,

$$E(P_r(y, \mathbf{G})) \leq \frac{b_y exp(\frac{2mr}{\frac{\tau}{2} * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N))})}{1 - \frac{S}{\mathbf{a}\cdot\mathbf{b}}}(\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{r-1} exp(p_{max}\frac{\eta(\frac{S}{\mathbf{a}\cdot\mathbf{b}})^2}{1 - \frac{S}{\mathbf{a}\cdot\mathbf{b}}(\eta)^{\frac{2}{\tau * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)}}}).$$

Proof. The proof is nearly identical to Theorem 19. Similar to equation (4.7), we have that

$$E(P_r^L(y,\mathbf{G})) \le \tag{4.73}$$

$$b_y exp(\frac{2mr}{\frac{\tau}{2} * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N))}) \sum_{\substack{k_1 \ge 1 \\ k_1+\sum_{i=2}^{r}(i-1)k_i = r}} \binom{\sum k_i}{k_1, ..., k_r} (\eta)^{1+\frac{2(i-1)k_i}{\tau * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)}} (\frac{\mathbf{a}\cdot\mathbf{b}}{S})^{k_1-1} \Pi_{i\ge 2} (\frac{p_{max}S}{\mathbf{a}\cdot\mathbf{b}})^{k_i}$$

where $\eta = log_{\tau\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)^2(m+1)^2$ and $r = \frac{m\tau}{2} * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)$. P_r^L is the number of paths starting from node y of length r, where require the last edge in the path to not be a repeating edge, $G \in \mathbf{G}$, and the k_i for $i > 1$ denote the number of repeated node blocks of length $i-1$. The key difference is instead of having $r^{(i-1)k_i}$ possible choices for each of the k_i repeating node blocks of length $i-1$, we claim that we have at most $([\tau * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)]^2(m+1)^2)^{1+\frac{2(i-1)}{\tau * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)}}$ choices for each repeating node block of length $i-1$. To see this note that by assumption, we can only have at most one cycle on any path of length bounded by $\frac{\tau}{2} * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)$. Since there at most $m+1$ node repetitions in the entire path of length r by construction, we claim that any repeating node block of length $i-1 \le \frac{\tau}{2} * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)$ has at most $[\tau * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)]^2(m+1)^2$ choices for the node block repetition.

In particular, we can have at most one unique (simple) cycle in the entire $\frac{\tau}{2} * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)$ portion of the path by assumption. There are at most $m+1$ choices for a particular cycle since there are only $m+1$ node repetitions in the path of length r. Furthermore, we can choose to place the cycle in at most i locations in a repeating node block of length i. In addition, we can concatenate the same cycle at most i times. This yields $i^2(m+1)$ possible choices for a repeating edge block of length i bounded by $\frac{\tau}{2} * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)$. For now we only consider the $(m+1)$ choices for the initial node in the repeating node block. Note that $i \le \tau * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)$ and we conclude that the number of possible choices for a repeating node block is bounded by $[\tau * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)]^2(m+1)^2$.

To apply a similar bound to a node block of any length, we divide the node block into smaller node blocks each of length $\frac{\tau}{2} * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)$ and possibly one node block that is smaller than the others. The number of choices for this larger repeating node block is then bounded by $([\tau * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)]^2(m+1))^{1+\frac{2(i-1)}{\tau * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)}}$ by considering the number of possible choices for each smaller repeating node block of length bounded by $\frac{\tau}{2} * log_{\frac{\mathbf{a}\cdot\mathbf{b}}{S}}(N)$.

We did not consider the number of choices for the repeating node blocks that occur aside from the appearance of the cycle and the first node in each repeating node block; we consider this issue at this juncture. One can show that for every portion of the path of length $\frac{\tau}{2} * log_{\frac{\mathbf{a} \cdot \mathbf{b}}{S}}(N)$, we need to multiply $[\tau * log_{\frac{\mathbf{a} \cdot \mathbf{b}}{S}}(N)]^2(m+1)^2$ by $exp(2m)$. More precisely, the number of choices for nodes in the repeating node blocks comes from the number of paths of fixed length from an arbitrary node that do not have node repetitions in the auxiliary graph formed by the edges we already visited in constructing a given path. More succinctly, we can construct our auxiliary graph by fixing the number of nodes to be, N_*, and having each node possess one incoming and one outgoing edge. Then we can add up to m additional edges to the auxiliary graph based on the upperbound of the m node repetitions in the path. Consequently, we can define $f_d(m)$ to be an upperbound for the number of paths without node repetitions of length d from any fixed node in this auxiliary graph constructed with m additional edges. Suppose the initial node that maximizes this quantity has out-degree $x_1 + 1$. It follows that

$$f_d(m) \leq \max_{x_1 \leq m, x_1 \in \mathbb{Z}} \sum_{i=1}^{x_1+1} f_{d-1}(m - x_1) = \max_{x_1 \leq m, x_1 \in \mathbb{Z}} (1 + x_1) f_{d-1}(m - x_1),$$

where the first inequality comes from the fact that since we are only considering paths without node repetitions. More specifically by considering the particular node, with out-degree $1 + x_1$, and graph that maximizes $f_d(m)$, we can instead bound this quantity by the number of paths of length $d - 1$ coming from this node's neighbors, where we essentially delete the node as we do not consider paths with node repetitions. Note that if our auxiliary graph with no additional edges, then we can let $f_{d_*}(0) = 1$ for any d_*. Furthermore, since the maximum number of paths of length 0 from a given node is bounded by 1, we get that $f_0(m_*) = 1$ for any m_*. Proceeding inductvely we conclude that,

$$f_d(m) \leq \max_{\substack{\sum_{i=1}^d x_i \leq m \\ x_i \geq 0, i \in \{1, \ldots, d\}}} \Pi_{i=1}^d (1 + x_i) \leq (1 + \frac{m}{d})^d \leq exp(m),$$

where the second to last inequality can be proven by induction on d. As we consider paths from the auxiliary graph *before and after* the appearance of the concatenation of cycles in the node block, we get a factor of $exp(2m)$. Since we have $\frac{r}{\frac{\tau}{2} * log_{\frac{\mathbf{a} \cdot \mathbf{b}}{S}}(N)}$ such portions in our path, this yields a total of $exp(\frac{2mr}{\frac{\tau}{2} * log_{\frac{\mathbf{a} \cdot \mathbf{b}}{S}}(N))})$ possible choices. Multiplying (4.73) by this quantity accounts for such choices in all possible repeating node blocks that may exist in our path. The proof then proceeds identically to that of Theorem 19 and we omit the details.

□

4.3 PARTITIONED CHUNG-LU MODEL

We now seek to extend our results to a generalization of the Chung-Lu model. Similar models have been defined in context to community detection for undirected graphs [21, 22, 48, 60, 65]. These models are special cases of the model below, where we construct our model in a general sense such that analogous proof statements still hold.

Definition 10. *We define the Chung-Lu Partitioned Random Graph Model such that we are given a collection of expected degree sequences for submatrices of our adjacency matrix. We construct an edge from node i to j by means of an independent Bernoulli random variable p_{ij} where p_{ij} is proportional to the product of the expected out-degree of node i and expected in-degree of node j of the corresponding submatrix.*

The first result we prove holds in considerable generality. Therefore, we introduce the following (more general) definition.

Definition 11. *We define the K-Partitioned Random Graph Model such that we are given a collection of expected degree sequences for submatrices of our adjacency matrix. We assign each node to one of K groups (or communities), denoted by the function $G(\cdot)$. We then construct an edge from node i to j by means of an independent Bernoulli random variable p_{ij} where p_{ij} depends on $G(i)$ and $G(j)$.*

We will find that the following norm will be helpful in proving bounds for the dominating eigenvalue in the Chung-Lu Partitioned Random Graph model. For completeness, we provide the following definition.

Definition 12. *Consider a vector $\mathbf{x} \in \mathbb{R}^{N\times 1}$. Denote $|\mathbf{x}|$ as the l1 norm (or taxicab norm for the vector). That is $|\mathbf{x}| = \sum_{i=1}^{N} |x_i|$. Furthermore for a matrix $B \in \mathbb{R}^{N\times N}$, we can also define $|B|$ to be the l1 norm of the matrix where $|B| = \sum_{i,j} |b_{ij}|$.*

For simplicity we will consider the case where there are two communities (analogous results hold when there are more than two communities).

Unsurprisingly, computing the number of paths and cycles becomes much more challenging when we incorporate partitions (communities) into our random graph model. We therefore introduce the following lemma that will facilitate the computation of otherwise unweildy expressions.

Lemma 17. *Consider the 2-Partitioned Random Graph Model. Denote the number of paths from node i_0 to node i_r of length r as $P_r[i_0 \to i_r]$. and define $p_{ij}(x,y)$ to be p_{ij} if node i in group x and node j in group y and 0 otherwise. Then,*

$$E(P_r[i_0 \to i_r]) = |\sum_{i_1,\ldots,i_{r-1}} [\Pi_{k=1}^{r-1} \mathbf{A}(i_k, i_{k+1}, \mathbf{i}_k)]\mathbf{p}| \tag{4.74}$$

where $|\cdot|$ denotes the taxicab norm, $\mathbf{i}_{k-1} = [(i_1,i_2),(i_2,i_3),\ldots,(i_{k-1},i_k)]$

and if $(i,j) \notin \mathbf{i}$ then,

$$\mathbf{A}(i,j,\mathbf{i}) = \begin{pmatrix} p_{ij}(1,1) & p_{ij}(1,1) & 0 & 0 \\ 0 & 0 & p_{ij}(2,1) & p_{ij}(2,1) \\ p_{ij}(1,2) & p_{ij}(1,2) & 0 & 0 \\ 0 & 0 & p_{ij}(2,2) & p_{ij}(2,2) \end{pmatrix}$$

else if $(i,j) \in \mathbf{i}$,

$$\mathbf{A}(i,j,\mathbf{i}) = \mathbf{G}_{ij} = \begin{pmatrix} G_{ij}(1,1) & G_{ij}(1,1) & 0 & 0 \\ 0 & 0 & G_{ij}(2,1) & G_{ij}(2,1) \\ G_{ij}(1,2) & G_{ij}(1,2) & 0 & 0 \\ 0 & 0 & G_{ij}(2,2) & G_{ij}(2,2) \end{pmatrix}$$

where $G_{ij}(c,d) = 1$ if node i is in group c and node j is in group d and $G_{ij}(c,d) = 0$ otherwise.

and

$$\mathbf{p} = [p_{i_0 i_1}(1,1), p_{i_0 i_1}(2,1), p_{i_0 i_1}(1,2), p_{i_0 i_1}(2,2)]^T$$

Furthermore for arbitrary i_0, i_1, i_2, consider the rth component of $\mathbf{A}(i_1,i_2)\mathbf{p}$.

- *If $r = 1$ this corresponds to the probability of a path under the constraint that $G(i_2) = 1, G(i_1) = 1$.*
- *If $r = 2$ this corresponds to the probability of a path under the constraint that $G(i_2) = 1, G(i_1) = 2$.*

- *If $r = 3$ this corresponds to the probability of a path under the constraint that $G(i_2) = 2, G(i_1) = 1$.*
- *If $r = 4$ this corresponds to the probability of a path under the constraint that $G(i_2) = 2, G(i_1) = 2$.*

Proof. We proceed by induction starting with $r = 2$.

So consider

$$\begin{pmatrix} p_{i_1i_2}(1,1) & p_{i_1i_2}(1,1) & 0 & 0 \\ 0 & 0 & p_{i_1i_2}(2,1) & p_{i_1i_2}(2,1) \\ p_{i_1i_2}(1,2) & p_{i_1i_2}(1,2) & 0 & 0 \\ 0 & 0 & p_{i_1i_2}(2,2) & p_{i_1i_2}(2,2) \end{pmatrix} \begin{pmatrix} p_{i_0i_1}(1,1) \\ p_{i_0i_1}(2,1) \\ p_{i_0i_1}(1,2) \\ p_{i_0i_1}(2,2) \end{pmatrix} =$$

$$= \begin{pmatrix} p_{i_0i_1}(1,1)p_{i_1i_2}(1,1) + p_{i_0i_1}(2,1)p_{i_1i_2}(1,1) \\ p_{i_0i_1}(1,2)p_{i_1i_2}(2,1) + p_{i_0i_1}(2,2)p_{i_1i_2}(2,1) \\ p_{i_0i_1}(1,1)p_{i_1i_2}(1,2) + p_{i_0i_1}(2,1)p_{i_1i_2}(1,2) \\ p_{i_0i_1}(1,2)p_{i_1i_2}(2,2) + p_{i_0i_1}(2,2)p_{i_1i_2}(2,2) \end{pmatrix}$$

Each entry of the vector classifies the probability of a path from i_0 to i_2 under different assumptions. It then follows from the definition that the first entry assesses the probability of a path where i_2 is in group 1 and i_1 is in group 1. The second entry details paths where i_2 is group 1 and i_1 is in group 2. The third entry specifies paths where i_2 is in group 2 and i_1 is in group 1. And finally, the fourth entry specifies where i_2 is in group 2 and i_1 is in group 2. In the case $(i_0, i_1) = (i_1, i_2)$, then the existence of the path does not depend on the edge (i_1, i_2) and we instead multiply the vector $\mathbf{p}$ by the matrix $\mathbf{G}_{i_1i_2}$ (as defined in the statement of Lemma 17. It then follows by linearity that taking the taxicab norm of the sum of such vectors (where we sum over all possible choces of i_1) will be the expected number of paths from i_0 to i_2.

Inductive Step: Suppose we were given a vector with the probability of the existence of a paths of length k (consisting of nodes $i_0, ..., i_k$) where each of the four components of the vector denote the probability where i_k is in group y and i_{k-1} is in group x. We symbolically denote this quantity as $p_k(x,y)$. Now to compute the probability of the existence of a path of length $k+1$, we then consider

$$\begin{pmatrix} p_{i_k i_{k+1}}(1,1) & p_{i_k i_{k+1}}(1,1) & 0 & 0 \\ 0 & 0 & p_{i_k i_{k+1}}(2,1) & p_{i_k i_{k+1}}(2,1) \\ p_{i_k i_{k+1}}(1,2) & p_{i_k i_{k+1}}(1,2) & 0 & 0 \\ 0 & 0 & p_{i_k i_{k+1}}(2,2) & p_{i_k i_{k+1}}(2,2) \end{pmatrix} \begin{pmatrix} p_k(1,1) \\ p_k(2,1) \\ p_k(1,2) \\ p_k(2,2) \end{pmatrix}$$

The output of which will yield a vector with four entries each considering different cases based on the group membership of the node i_k and i_{k+1}. Alternatively, if the edge (i_k, i_{k+1}) has been already visited earlier in the path, we multiply the vector by the matrix $\mathbf{G}_{i_k i_{k+1}}$. It then follows by linearity that taking the taxicab norm of the sum of such probabilities of existence of paths of length k (over all possible choces of $i_1, ..., i_k$) will be the expected number of paths from i_0 to i_{k+1}.

□

To bound the expected number of paths of length r under this partitioned random graph model, we will want to express the bounds of the norm of a matrix vector product in terms of the dominating eigenvalue of the matrix. As such, we need the following result (see [44], page 494). Since the idea behind their proof is quite succinct, for completeness, we include their proof in this work.

Lemma 18. *Let $B \in \mathbb{R}^{N\times N}$ be a (entry-wise) nonnegative matrix and let x be the eigenvector corresponding to the dominating eigenvalue. Assume that the eigenvector x is strictly positive. Furthermore let $b_{ij}^{(m)}$ denote the i,jth entry of B^m. Then for all integers m and integers j such that $1 \leq j \leq N$, we have that*

$$\sum_{i=1}^{n} b_{ij}^{(m)} \leq \frac{\max_k x_k}{\min_k x_k} \rho(B)^m$$

and

$$\frac{\min_k x_k}{\max_k x_k} \rho(B)^m \leq \sum_{j=1}^{n} b_{ij}^{(m)}$$

Proof. Consider the eigenvector corresponding to the dominating eigenvalue $\lambda_{max} = \rho(B)$ and its eigenvector x. Then we have that for any j,

$$\min_k x_k \sum_i b_{ij}^{(m)} \leq \sum_i b_{ij}^{(m)} x_i = [B^m x]_j = \rho(B)^m x_j \leq \rho(B)^m \max_j x_j$$

$$\implies \sum_i b_{ij}^{(m)} \leq \rho(B)^m \frac{\max_k x_k}{\min_k x_k}$$

□

With the above Lemma at hand, we need to bound $\max_k x_k$ and $\min_k x_k$ for the dominating eigenvector.

Lemma 19. *Let r_{max}, c_{max} be the maximum row sum and column sum of $B \in \mathbb{R}^{n\times n}$. In addition, suppose every entry is at least equal to $m > 0$. Denote $\rho(B) = \lambda_{max}$. Since by the Gresgorin Disc Theorem $\lambda_{max} \leq \min(r_{max}, c_{max})$. Then*

$$\frac{m}{\min(c_{max}, r_{max}) - m(n-1)} \leq \frac{m}{\lambda_{max} - m(n-1)} \leq \frac{\min_k x_k}{\max_k x_k}$$

and

$$\frac{\max_k x_k}{\min_k x_k} \leq \frac{\lambda_{max} - m(n-1)}{m} \leq \frac{\min(c_{max}, r_{max}) - m(n-1)}{m}$$

Proof. Consider the eigenvector x and require that $\sum_{j=1}^n x_j = 1$ where we are guaranteed that each entry in the eigenvector is non-negative by the Perron-Frobenius Theorem. Then we have for all k,

$$m = \sum_{j=1}^n m x_j \leq \sum_{j=1}^n b_{jk} x_j = \lambda_{max} x_k$$

It then follows that for all k

$$\frac{m}{\lambda_{max}} \leq x_k$$

Consequently,

$$\frac{m}{\lambda_{max}} \leq \min_k x_k$$

. Furthermore, since $\sum_j x_j = 1$, we have that

$$\max_k x_k \leq 1 - \frac{m * (n-1)}{\lambda_{max}}$$

This implies that

$$\frac{m}{\lambda_{max} - m(n-1)} \leq \frac{\min_k x_k}{\max_k x_k}$$

and

$$\frac{\max_k x_k}{\min_k x_k} \leq \frac{\lambda_{max} - m(n-1)}{m}$$

□

We then get the immediate corollary,

Corollary 16. *Let B be an (entry-wise) non-negative matrix, B^2 be an entry-wise positive matrix and let x be the eigenvector corresponding to the dominating eigenvalue. Furthermore let $b_{ij}^{(m)}$ denote the i,jth entry of B^m .Let c_{max} be the maximum row sum of $B^2 \in \mathbb{R}^{n\times n}$ and furthermore suppose every entry is at least equal to 1, (hence $r_{max} \geq n$). Then*

$$[\sum_{j=1}^{n} b_{ij}^{(m)}]^{\frac{1}{m}} \leq {c_{max}}^{\frac{1}{m}} \rho(B)$$

and

$$(c_{max})^{-\frac{1}{m}} \rho(B) \leq [\sum_{j=1}^{n} b_{ij}^{(m)}]^{\frac{1}{m}}$$

We now prove our desired result regarding the expected number of paths of length r.

Theorem 25. *Consider a realization of a graph in the 2-Partitioned Chung-Lu random graph model. Denote P_r as the number of paths of length r. Define*

$$\mathbf{P} = \begin{pmatrix} \frac{\mathbf{a}^{(11)} \cdot \mathbf{b}^{(11)}}{S_{11}} & \frac{\mathbf{a}^{(21)} \cdot \mathbf{b}^{(11)}}{S_{21}} & 0 & 0 \\ 0 & 0 & \frac{\mathbf{a}^{(12)} \cdot \mathbf{b}^{(21)}}{S_{12}} & \frac{\mathbf{a}^{(22)} \cdot \mathbf{b}^{(21)}}{S_{22}} \\ \frac{\mathbf{a}^{(11)} \cdot \mathbf{b}^{(12)}}{S_{11}} & \frac{\mathbf{a}^{(21)} \cdot \mathbf{b}^{(12)}}{S_{21}} & 0 & 0 \\ 0 & 0 & \frac{\mathbf{a}^{(12)} \cdot \mathbf{b}^{(22)}}{S_{12}} & \frac{\mathbf{a}^{(22)} \cdot \mathbf{b}^{(22)}}{S_{22}} \end{pmatrix}$$

.

Furthermore suppose for all choices of m and $i_1, ..., i_{m+1}$, that there exists an α such that

$$\begin{pmatrix} b_{max}^{(1,1)} & b_{max}^{(1,1)} & 0 & 0 \\ 0 & 0 & b_{max}^{(2,1)} & b_{max}^{(2,1)} \\ b_{max}^{(1,2)} & b_{max}^{(1,2)} & 0 & 0 \\ 0 & 0 & b_{max}^{(2,2)} & b_{max}^{(2,2)} \end{pmatrix} \Pi_{k=1}^{m} \mathbf{G}_{i_k i_{k+1}} \begin{pmatrix} a_{max}^{(1,1)}/S_{11} & 0 & 0 & 0 \\ 0 & a_{max}^{(2,1)}/S_{21} & 0 & 0 \\ 0 & 0 & a_{max}^{(1,2)}/S_{12} & 0 \\ 0 & 0 & 0 & a_{max}^{(2,2)}/S_{22} \end{pmatrix} \leq \alpha \mathbf{P} \qquad (4.75)$$

where $\mathbf{G}_{i_k i_{k+1}}$ *is defined in Lemma 17 and the inequality holds entry-wise. Furthermore assume that* $\rho(\mathbf{P}) > 2$*.*

Then

$$E(P_r) \leq 8 * S * c_{max} * \rho(\mathbf{P})^{r-1} * exp(\frac{r^2\alpha\rho(\mathbf{P})^{-1}}{1 - r\rho(\mathbf{P})^{-1}})$$

where c_{max} *is the maximum column sum of* $\mathbf{P}^2$ *and the vectors* $\mathbf{a}^{(x,y)}, \mathbf{b}^{(x,y)}$ *are the corresponding expected row and column sums of the partitions in* $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ *as defined in the Partitioned Chung-Lu random graph model.*

Remark. While condition (4.75) at first look may appear like a difficult condition to satisfy, this in fact is not so. Upon careful observation of Lemma 17, $\mathbf{G}_{ij}$ consists of two columns that are from the standard unit basis and two columns that are zero. Consequently when we perform a matrix vector multiplication, the taxicab norm resulting vector cannot be greater than the taxicab norm of the input. Equivalently, we can write for any vector $\mathbf{v}$, that $|\mathbf{G}_{ij}\mathbf{v}| \leq |\mathbf{v}|$. Now a matrix product of the form (4.75) indicates a repeating edge block of length m. If $\rho(\mathbf{P}) \to \infty$, then since $\mathbf{G}$ has the property that $|\mathbf{G}_{ij}\mathbf{v}| \leq |\mathbf{v}|$, we can often satisfy condition (4.75) with ease. It is worth mentioning that even if we cannot satisfy (4.75), we could still prove a useful generalization of Theorem 25 by requiring that for each m we can find an α_m such that

$$\begin{pmatrix} b_{max}^{(1,1)} & b_{max}^{(1,1)} & 0 & 0 \\ 0 & 0 & b_{max}^{(2,1)} & b_{max}^{(2,1)} \\ b_{max}^{(1,2)} & b_{max}^{(1,2)} & 0 & 0 \\ 0 & 0 & b_{max}^{(2,2)} & b_{max}^{(2,2)} \end{pmatrix} \Pi_{k=1}^{m} \mathbf{G}_{i_k i_{k+1}} \begin{pmatrix} a_{max}^{(1,1)}/S_{11} & 0 & 0 & 0 \\ 0 & a_{max}^{(2,1)}/S_{21} & 0 & 0 \\ 0 & 0 & a_{max}^{(1,2)}/S_{12} & 0 \\ 0 & 0 & 0 & a_{max}^{(2,2)}/S_{22} \end{pmatrix} \leq \alpha_m \mathbf{P}^m. \quad (4.76)$$

Alternatively, we can also satisfy (4.75), if the product of the norms of the left and right matrices on the left hand side of (4.75), are sufficiently small, analogous to the case where $p_{max} \to 0$. With this in mind, we now provide the proof.

Proof. The proof is similar to Theorem 19. We first try to construct an upperbound to the expected number of paths of length r where there are no repeating edges. Denote this quantity by $P_{r,0}$.

By Lemma 17, we find that

$$E(P_{r,0}) \leq$$

$$|\sum_{\substack{i_0\neq\\ \ldots\neq i_r}} \Pi_{k=1}^{r-1} \begin{pmatrix} p_{i_k i_{k+1}}(1,1) & p_{i_k i_{k+1}}(1,1) & 0 & 0 \\ 0 & 0 & p_{i_k i_{k+1}}(2,1) & p_{i_k i_{k+1}}(2,1) \\ p_{i_k i_{k+1}}(1,2) & p_{i_k i_{k+1}}(1,2) & 0 & 0 \\ 0 & 0 & p_{i_k i_{k+1}}(2,2) & p_{i_k i_{k+1}}(2,2) \end{pmatrix} \begin{pmatrix} p_{i_0 i_1}(1,1) \\ p_{i_0 i_1}(2,1) \\ p_{i_0 i_1}(1,2) \\ p_{i_0 i_1}(2,2) \end{pmatrix}| \tag{4.77}$$

Recall that $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$. We define $a_{i_k}^{(x,y)}$ to be 0 if i_k does not belong to group x. If i_k does belong to group x, then $a_{i_k}^{(x,y)}$ will be the expected row sum corresponding to node i_k in the submatrix A_{yx}. Analogously, we define $b_{i_k}^{(x,y)}$ to be 0 if i_k does not belong to group y. If i_k does belong to group y, then $b_{i_k}^{(x,y)}$ will be the expected column sum corresponding to node i_k in the submatrix A_{yx}. Consequently by the definition of the Chung-Lu random graph, we have that

$$\begin{pmatrix} p_{i_k i_{k+1}}(1,1) & p_{i_k i_{k+1}}(1,1) & 0 & 0 \\ 0 & 0 & p_{i_k i_{k+1}}(2,1) & p_{i_k i_{k+1}}(2,1) \\ p_{i_k i_{k+1}}(1,2) & p_{i_k i_{k+1}}(1,2) & 0 & 0 \\ 0 & 0 & p_{i_k i_{k+1}}(2,2) & p_{i_k i_{k+1}}(2,2) \end{pmatrix} =$$
$$\begin{pmatrix} a_{i_{k+1}}^{(1,1)}/S_{11} & 0 & 0 & 0 \\ 0 & a_{i_{k+1}}^{(2,1)}/S_{21} & 0 & 0 \\ 0 & 0 & a_{i_{k+1}}^{(1,2)}/S_{12} & 0 \\ 0 & 0 & 0 & a_{i_{k+1}}^{(2,2)}/S_{22} \end{pmatrix} \begin{pmatrix} b_{i_k}^{(1,1)} & b_{i_k}^{(1,1)} & 0 & 0 \\ 0 & 0 & b_{i_k}^{(2,1)} & b_{i_k}^{(2,1)} \\ b_{i_k}^{(1,2)} & b_{i_k}^{(1,2)} & 0 & 0 \\ 0 & 0 & b_{i_k}^{(2,2)} & b_{i_k}^{(2,2)} \end{pmatrix} \tag{4.78}$$

Because this is rather unweildy, we will denote

$$\mathbf{A}_{i_{k+1}} = \begin{pmatrix} a_{i_{k+1}}^{(1,1)}/S_{11} & 0 & 0 & 0 \\ 0 & a_{i_{k+1}}^{(2,1)}/S_{21} & 0 & 0 \\ 0 & 0 & a_{i_{k+1}}^{(1,2)}/S_{12} & 0 \\ 0 & 0 & 0 & a_{i_{k+1}}^{(2,2)}/S_{22} \end{pmatrix} \tag{4.79}$$

$$\mathbf{B}_{i_k} = \begin{pmatrix} b_{i_k}^{(1,1)} & b_{i_k}^{(1,1)} & 0 & 0 \\ 0 & 0 & b_{i_k}^{(2,1)} & b_{i_k}^{(2,1)} \\ b_{i_k}^{(1,2)} & b_{i_k}^{(1,2)} & 0 & 0 \\ 0 & 0 & b_{i_k}^{(2,2)} & b_{i_k}^{(2,2)} \end{pmatrix} \tag{4.80}$$

Now it follows from equation (4.77), (4.79) and (4.80),

$$E(P_{r,0}) \leq |\sum_{\substack{i_0 \neq \\ \dots \neq i_r}} \Pi_{k=1}^{r-1} \mathbf{A}_{i_{k+1}} \mathbf{B}_{i_k} \begin{pmatrix} p_{i_0 i_1}(1,1) \\ p_{i_0 i_1}(2,1) \\ p_{i_0 i_1}(1,2) \\ p_{i_0 i_1}(2,2) \end{pmatrix} | \tag{4.81}$$

Which we can rewrite as,

$$E(P_{r,0}) \leq |\sum_{i_0 \neq \dots \neq i_r} \mathbf{A}_{i_r} \Pi_{k=1}^{r-1} \mathbf{B}_{i_k} \mathbf{A}_{i_k} \begin{pmatrix} b_{i_0}^{(1,1)} \\ b_{i_0}^{(2,1)} \\ b_{i_0}^{(1,2)} \\ b_{i_0}^{(2,2)} \end{pmatrix} | \tag{4.82}$$

Recall from (4.79) and (4.80) that

$$\mathbf{B}_{i_k} \mathbf{A}_{i_k} = \begin{pmatrix} b_{i_k}^{(1,1)} a_{i_k}^{(1,1)}/S_{11} & b_{i_k}^{(1,1)} a_{i_k}^{(2,1)}/S_{21} & 0 & 0 \\ 0 & 0 & b_{i_k}^{(2,1)} a_{i_k}^{(1,2)}/S_{12} & b_{i_k}^{(2,1)} a_{i_k}^{(2,2)}/S_{22} \\ b_{i_k}^{(1,2)} a_{i_k}^{(1,1)}/S_{11} & b_{i_k}^{(1,2)} a_{i_k}^{(2,1)}/S_{21} & 0 & 0 \\ 0 & 0 & b_{i_k}^{(2,2)} a_{i_k}^{(1,2)}/S_{12} & b_{i_k}^{(2,2)} a_{i_k}^{(2,2)}/S_{22} \end{pmatrix} \tag{4.83}$$

And from the statement of this Theorem, Theorem 25, recall the definition of $\mathbf{P}$. It follows that by summing over all possible choices of nodes for i_k that,

$$\mathbf{P} = \sum_{i_k=1}^{N} \mathbf{B}_{i_k} \mathbf{A}_{i_k} \tag{4.84}$$

Hence we conclude from (4.82) that,

$$E(P_{r,0}) \leq |\sum_{i_0,\dots,i_r} \mathbf{A}_{i_r} \Pi_{k=1}^{r-1} \mathbf{B}_{i_k} \mathbf{A}_{i_k} \begin{pmatrix} b_{i_0}^{(1,1)} \\ b_{i_0}^{(2,1)} \\ b_{i_0}^{(1,2)} \\ b_{i_0}^{(2,2)} \end{pmatrix} | = |\mathbf{I} * \mathbf{P}^{r-1} \begin{pmatrix} S_{11} \\ S_{12} \\ S_{21} \\ S_{22} \end{pmatrix} | \tag{4.85}$$

where S_{ij} denotes the expected sum of entries in the submatrix A_{ij}. Now by definition of $\mathbf{P}$, $\mathbf{P}^4$ is an entry-wise positive matrix where each entry is bounded below by 1. Define c_{max} to be the maximum column sum of $\mathbf{P}^2$. It follows by Corollary 16, that each column sum of $\mathbf{P}^{r-1}$ is bounded above by $c_{max}\rho(\mathbf{P}^{r-1})$. Hence we conclude that

$$E(P_{r,0}) \leq 4c_{max}(S_{11} + S_{12} + S_{21} + S_{22})\rho(\mathbf{P}^{r-1})$$

We can simplify this by defining $S_{11} + S_{12} + S_{21} + S_{22} = S$ and get that,

$$E(P_{r,0}) \leq 4c_{max}S\rho(\mathbf{P})^{r-1}$$

Now we repeat the argument utilized in Theorem 20 counting paths with duplicate edges.

We will compute the bound when there is precisely one duplicate edge (where the duplicate edge is not in the end).

Suppose that the repeating edge in the path is the mth edge.

Then we have that

$$E(P_{r,1,m}) \leq |\sum_{only\ edge\ m\ repeats} \Pi_{k=1,k\neq m}^{r-1} \begin{pmatrix} p_{i_k i_{k+1}}(1,1) & p_{i_k i_{k+1}}(1,1) & 0 & 0 \\ 0 & 0 & p_{i_k i_{k+1}}(2,1) & p_{i_k i_{k+1}}(2,1) \\ p_{i_k i_{k+1}}(1,2) & p_{i_k i_{k+1}}(1,2) & 0 & 0 \\ 0 & 0 & p_{i_k i_{k+1}}(2,2) & p_{i_k i_{k+1}}(2,2) \end{pmatrix} \begin{pmatrix} p_{i_0 i_1}(1,1) \\ p_{i_0 i_1}(2,1) \\ p_{i_0 i_1}(1,2) \\ p_{i_0 i_1}(2,2) \end{pmatrix} |$$

Using the argument as before (4.78),(4.79),(4.80) we get that

$$E(P_{r,1,m}) \leq |\mathbf{P}^{r-m-1} \begin{pmatrix} b_{max}^{(1,1)} & b_{max}^{(1,1)} & 0 & 0 \\ 0 & 0 & b_{max}^{(2,1)} & b_{max}^{(2,1)} \\ b_{max}^{(1,2)} & b_{max}^{(1,2)} & 0 & 0 \\ 0 & 0 & b_{max}^{(2,2)} & b_{max}^{(2,2)} \end{pmatrix} \mathbf{G}_{m,m+1} \begin{pmatrix} a_{max}^{(1,1)}/S_{11} & 0 & 0 & 0 \\ 0 & a_{max}^{(2,1)}/S_{21} & 0 & 0 \\ 0 & 0 & a_{max}^{(1,2)}/S_{12} & 0 \\ 0 & 0 & 0 & a_{max}^{(2,2)}/S_{22} \end{pmatrix} \mathbf{P}^{m-2}\mathbf{S}|$$

where $\mathbf{S} = (S_{11}, S_{12}, S_{21}, S_{22})$

But by assumption 4.75, we conclude that

$$E(P_{r,1,m}) \leq |\mathbf{P}^{r-m-1}\alpha\mathbf{P}\mathbf{P}^{m-2}\mathbf{S}| \leq \alpha|\mathbf{P}^{r-2}\mathbf{S}| \leq 4Sc_{max}\alpha\rho(\mathbf{P})^{r-2}$$

Denote the number of expected paths of length r where the last edge just not repeat as $E(P_r^L)$.

It follows using an anlogous argument from Theorem 19 we have the following inequality, where we denote the expected number of paths of length r where the last edge does not repeat $E(P_r^L)$ and k_i is the number of repeating edge blocks of length $i-1$,

$$E(P_r^L) \leq 4Sc_{max} \sum_{\substack{k_1+\sum_{i=2}^{r}(i-1)k_i=r \\ k_1\geq 1}} \binom{\sum k_i}{k_1, ..., k_r} [k_1^{r-k_1}]\rho(\mathbf{P})^{k_1-1}\Pi_{i\geq 2}(\alpha[\rho(\mathbf{P})^{-1}])^{k_i} \tag{4.86}$$

Then continuing with the argument in Theorem 19, it follows that,

$$E(P_r^L) \leq 4Sc_{max}\rho(\mathbf{P})^{r-1} \sum_{\substack{k_1+\sum_{i=2}^{r}(i-1)k_i=r \\ k_1\geq 1}} r^{ik_i}\rho(\mathbf{P})^{-(i-2)k_i}\Pi_{i\geq 2}\frac{(\alpha[\rho(\mathbf{P})^{-1}])^{k_i}}{k_i!} \tag{4.87}$$

$$E(P_r^L) \leq 4Sc_{max}\rho(\mathbf{P})^{r-1} \sum_{\substack{k_1+\sum_{i=2}^{r}(i-1)k_i=r \\ k_1\geq 1}} \Pi_{i\geq 2}\frac{(\alpha\rho(\mathbf{P})[r\rho(\mathbf{P})^{-1}]^i])^{k_i}}{k_i!} \tag{4.88}$$

$$E(P_r^L) \leq 4Sc_{max}\rho(\mathbf{P})^{r-1} \sum_{k_2=0,\ldots,k_r=0}^{\infty} \Pi_{i\geq 2}\frac{(\alpha\rho(\mathbf{P})[r\rho(\mathbf{P}^{-1})]^i)^{k_i}}{k_i!} \tag{4.89}$$

$$E(P_r^L) \leq 4Sc_{max}\rho(\mathbf{P})^{r-1}exp(\frac{\alpha[r^2\rho(\mathbf{P})^{-1}]}{1-r\rho(\mathbf{P})^{-1}}) \tag{4.90}$$

And we conclude that

$$E(P_r) \leq \sum_{i=1}^{r} E(P_i^L) \leq 8Sc_{max}\rho(\mathbf{P})^{r-1}exp(\frac{\alpha[r^2\rho(\mathbf{P})^{-1}]}{1-r\rho(\mathbf{P})^{-1}})$$

as $\rho(\mathbf{P}) > 2$.

□

We now have the following concentration result.

Corollary 17. *Consider a realization of a graph A in the 2-Partitioned Chung-Lu random graph model. Denote P_r as the number of paths of length r. Define*

$$\mathbf{P} = \begin{pmatrix} \frac{\mathbf{a}^{(11)}\cdot\mathbf{b}^{(11)}}{S_{11}} & \frac{\mathbf{a}^{(21)}\cdot\mathbf{b}^{(11)}}{S_{21}} & 0 & 0 \\ 0 & 0 & \frac{\mathbf{a}^{(12)}\cdot\mathbf{b}^{(21)}}{S_{12}} & \frac{\mathbf{a}^{(22)}\cdot\mathbf{b}^{(21)}}{S_{22}} \\ \frac{\mathbf{a}^{(11)}\cdot\mathbf{b}^{(12)}}{S_{11}} & \frac{\mathbf{a}^{(21)}\cdot\mathbf{b}^{(12)}}{S_{21}} & 0 & 0 \\ 0 & 0 & \frac{\mathbf{a}^{(12)}\cdot\mathbf{b}^{(22)}}{S_{12}} & \frac{\mathbf{a}^{(22)}\cdot\mathbf{b}^{(22)}}{S_{22}} \end{pmatrix}$$

.

Furthermore suppose that condition (4.75) from Theorem 25 holds.

Then for every $\epsilon > 0$, there exists δ_1,δ_2 such that if $\frac{1}{N} < \delta_1$ and $log(N)^2\rho(\mathbf{P})^{-1} < \delta_2$, then

$$Pr(\rho(A) \leq (1+\epsilon)\rho(\mathbf{P})) \geq 1-\epsilon.$$

Proof. The proof is analogous to Theorem 20. Invoking Theorem 25 where we consider paths of length $r = log(N)^2$, Lemma 10 and Markov's Inequality yield the result. $\square$

In a similar spirit to Corollary 17, we anticipate that a similar proof technique will yield analogous concentration results regarding the lower bound for the dominating eigenvalue in the case when $\rho(\mathbf{P}) \to \infty$. In addition, Lemma 16, the lemma that restrict the number of simple cycles that can appear on a path of a prescribed length, holds for an ensemble of random graph models, not just the Chung-Lu model. As such, we also expect that concentration results regarding the dominating eigenvalue also hold, similar to the case where $p_{max} \to 0$ in the prior section, where instead we demand that $\max_{x,y} \max \frac{\mathbf{a}^{(x,y)}}{S_{x,y}} \max_{x,y} \max \mathbf{b}^{(x,y)} \to 0$.

To further attest to the validity of Theorem 25 and Corollary 17, we considered five different expected partitioned degree sequences associated with a network containing 200, 400, 600, 800 and 1000 nodes. For each expected partitioned degree sequence, we generated 100 realizations from the 2-Partitioned Chung-Lu random graph model. Then we computed the dominating eigenvalue from each of these realizations. In Figure 4.3, we plotted box plots corresponding to the difference from the observed dominating eigenvalue and $\rho(\mathbf{P})$ in pink and the difference between the dominating eigenvalue and $\frac{\mathbf{a} \cdot \mathbf{b}}{S}$ in blue. As suggested by Figure 4.3, the prediction from $\frac{\mathbf{a} \cdot \mathbf{b}}{S}$ appears increasingly inaccurate as we increase the size of the network, in contrast to the predictor $\rho(\mathbf{P})$, which appears to improve as we consider increasingly large networks.

Determining the dominating eigenvalue of the adjacency matrix can have a profound effect on the underlying dynamics of the network. For example consider a susceptible-infected-susceptible (SIS) epidemiological model, where at each step an infected node infects a neighbor with probability $\beta \Delta t$ and recovers (from sick to healthy) with probability Δt, where Δt denotes the length of the time step. We then have the following result,

Theorem 26 (Ganesh, Massoulie, Towsley [39]). *Consider an SIS epidemiological model, where infected nodes infect neighbors with probability $\beta \Delta t$ at each time step and recover with probability Δt. Furthermore, suppose our adjacency matrix $A \in \mathbb{R}^{N \times N}$ is symmetric. Then for Δt sufficiently small, if $\rho(A) < \frac{1}{\beta}$, then the expected (stopping) time for the network to*

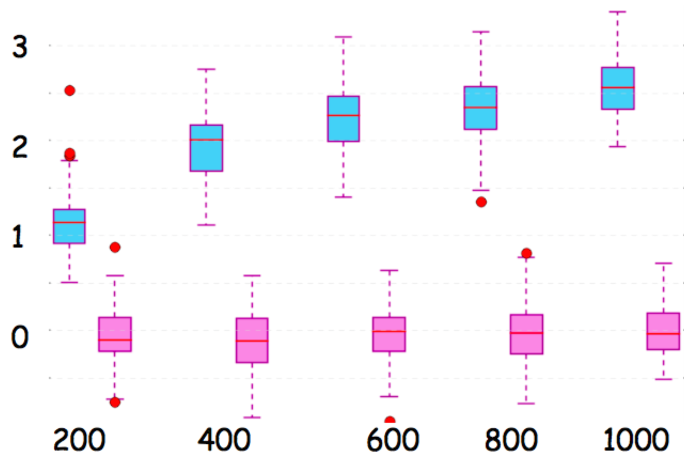

Figure 4.3.1: The x axis indicates networks of different sizes generated by the 2-Partitioned Chung-Lu model. The y axis denotes the difference between the dominating eigenvalue and the chosen predictor. The blue box and whisker plots use the predictor $\frac{\mathbf{a}\cdot\mathbf{b}}{S}$ and the pink box and whisker plots use the predictor $\rho(\mathbf{P})$.

be infection free from any initial condition is $O(log(N))$.

While our adjacency matrices are not symmetric, Theorem 26 relates the dominating eigenvalue of the adjacency matrix to the stability of the healthy state and provides a framework for constructing cases where differences in the spectral radius of the adjacency matrix between the Chung-Lu and Partitioned Chung-Lu model could have severe repercussions on the dynamics.

Consequently in Figure 4.3, we generated three realizations from the 2-Partitioned Chung-Lu models all with approximately the same value for $\frac{\mathbf{a}\cdot\mathbf{b}}{S}$, but different values for $\rho(\mathbf{P})$, and simulated 100 trials of the SIS epidemiolgocial stochastic process for each choice of $\beta \in \{.05, .06, .07\}$ with the initial condition that half of our network starts out infected. As expected $\rho(\mathbf{P})$ accurately predicted the network resilience to the pathogen; in contrast, the predictor $\frac{\mathbf{a}\cdot\mathbf{b}}{S}$, could not effectively discern differences among the networks.

As another example to further illustrate the importance of the spectral radius of the adjacency matrix on the dynamics of the network, we consider the Kuramoto model. We

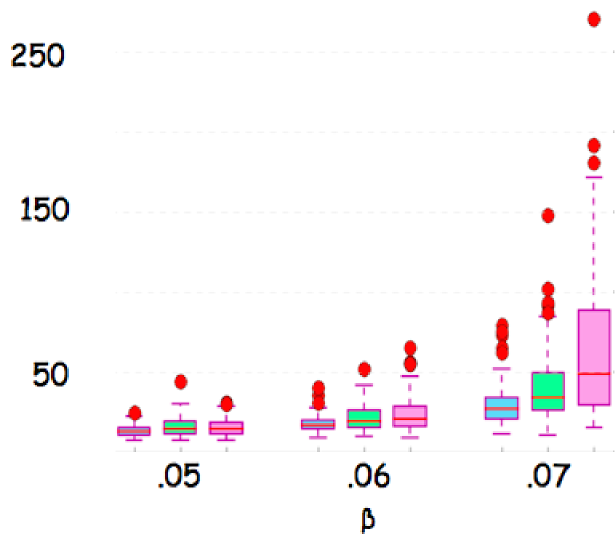

Figure 4.3.2: Distribution of Stopping Times in the SIS Model for Networks with Community Structure. The three colored box and whisker plots correspond to the three different networks, the x axis indicates the value for the paramater β and the y axis indicates the stopping time when nodes can no longer be infectious.

can define the Kuramoto model with and without noise. In particular if there is no noise, then the dynamics are described by the following system of differential equations where

$$\frac{d\theta_i}{dt} = \omega_i + K \sum_j A_{ij} \sin(\theta_i - \theta_j) \tag{4.91}$$

i identifies the node in our network, A is our adjacency matrix, ω_i is the 'natural frequency' of node i and K is a (non-negative) parameter indicating coupling strength. In terms of applications, the Kuramoto model is a popular toy model for considering the internactions of a biological neural network. In this context, we are often interested if the neurons exhibit synchronous spiking, which may be indiciative of schizophrenia, Parkinson's or Alzheimer's disease. Mathematically, we can construct a synchrony parameter, r ,by representing each θ_i as a vector on the unit circle. We then take the magnitude of the weighted sum of the vectors, where we weight the vector based on the out-degree of node i, and then normalize the magnitude so that the synchrony parameter can range only from 0 to 1. If the network behaves in a synchronous fashion then we expect the sum of such vectors to be very large and the synchrony parameter to be approximately 1. If in fact the vectors associated with the θ_i point in an ensemble of different directions, we expect cancellation and anticipate that r will be close to 0.

More succinctly we define our synchrony parameter r as,

$$r = \left\| \frac{\sum_{j=1,m=1}^{N} A_{jm} e^{i\theta_m}}{S} \right\|, \tag{4.92}$$

where S is the number of edges in our network and $\|\cdot\|$ is the euclidean norm. We now have the following result,

Theorem 27 (Restrepo, Ott and Hunt [72])**.** *Given an adjacency matrix $A \in \mathbb{R}^{N\times N}$, suppose that the natural frequences ω_i are randomly chosen, independent from node i. Furthermore suppose the ω_i follow a unimodal distribution $\Omega(\omega)$ which is symmetric about the maximum and the maximum value of $\Omega(\omega)$ occurs without loss of generality at $\omega = 0$. Then asymptotically as $N \to \infty$, the onset of solutions where the synchrony parameter $r > 0$, as defined in (4.92) for the Kuramoto model (4.91) starts when the coupling $K > \frac{2}{\pi\Omega(0)\rho(A)}$.*

Informally, the spectral radius determines the stability of the asynchronous state in the network. Consequently in Figure 4.3, we numerically tested the impact of the spectral radius of realizations from the 2-Partitioned Chung-Lu random graph model, where all three networks attain roughly the same value for $\frac{\mathbf{a} \cdot \mathbf{b}}{S}$, but have different values for $\rho(\mathbf{P})$. We then run 50 simulations of the Kuramoto dynamics for each network and record the distribution of the time average of the synchrony parameter r. As anticipated by Theorem 27, $\rho(\mathbf{P})$ correctly predicts which networks exhibit greater amounts of synchrony when K is not *too large*.

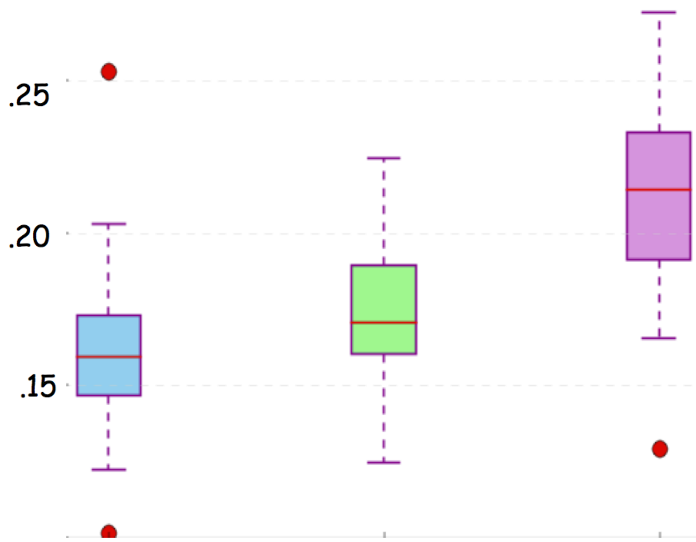

Figure 4.3.3: Distribution of the Synchrony Parameter in the Kuramoto Model with Noise. The three colored box and whisker plots correspond to the three different networks and the y axis marks the synchrony paramater r. Here, we added independent white gaussian noise to each node in the the Kuramoto model (4.91) and chose random initial conditions each of the different trials.

5.0 CONCLUSIONS

The challenge of identifying families of graphs that behave similarly under a prescribed dynamical process arises in numerous applications including ecology, epidemiology and neuroscience. More specifically, prior evidence has demonstrated that the degree sequence of a graph can impact the dynamical behavior of the network. In order to better understand the intricacies of the dependence of the degree sequence on the dynamics, we initially considered the Uniform Model, where we randomly construct a graph that realizes a given bi-degree sequence in an unbiased fashion. Since we want to numerically observe the impact of the degree sequence on the dynamics, in the first two chapters we devoted our efforts to build machinery that would assist us in constructing random realizations of networks in the Uniform Model. Initially in Chapter 2, we developed techniques to quickly determine if we can construct a graph from a given degree sequence. Employing our easily verifiable sufficient conditions for graphicality can not only assist us in quickly identifying graphic degree sequences but also can supplement existing techniques for randomly generating degree sequences by guaranteeing that the degree sequence output will be graphic as well. [23, 53].

To assist us in sampling graphs from the Uniform Model, in Chapter 3 we provided a novel method for constructing asymptotics up to arbitrary order for the number of graphs that realize a given degree sequence,extending the work of Greenhill, McKay and Wang [42]. More specifically, Greenhill, McKay and Wang[42] provide an asymptotic enumeration result when the maximum degree is $o(S^{\frac{1}{3}})$; in constrast, our asymptotic enumeration results apply to degree sequences allowing for the maximum degree to be $O(S^{\frac{1}{2}-\tau})$ where S is the sum of the number of edges and the positive variable τ depends on the degree sequence $\mathbf{d} = (\mathbf{a}, \mathbf{b})$. As many real world networks contain a handful of nodes of large degree, scaling larger than $o(S^{\frac{1}{3}})$, we expect that our results will be helpful for generating realizations of graphs in the uniform model with degree sequences similar to those observed in real world networks.

Our primary motivation for pursuing the asymptotic enumeriation problem dealt with constructing realizations of the uniform random graph model to perform numerical simulations to comprehend the impact of the degree sequence on the underlying dynamics; nevertheless, our asymptotic results have considerable theoretical significance as well. Identifying unique properties of different random graph models is important for selecting the appropriate random graph model for a given application. For this purpose, we can invoke our asymptotic enumeration results to find probabilities for the likelihood two nodes share an edge and compare these probabilities with other random graph models. We demonstrated in Section 3.7 that the probability that two nodes share an edge in the Uniform Model converges to the probability given in the Chung-Lu random graph model with a multiplicative error bounded above by $(1 + O(S^{-2\tau}))$.

Such conclusions regarding the similarity between the Uniform Model and the Chung-Lu random graph model should be taken with care. In particular, we stress that the speed of the convergence depends on the degree sequence and there are many degree sequences where the probabilities in the Uniform Model converge rather slowly to the probabilities in the Chung-Lu model. This concern has been numerically demonstrated in [76] where they considered the impact of different 'uniform' random graph models on the network architecture. The implications regarding the similarity (or lack thereof) of different 'uniform' models affects not only problems pertaining to stability of solutions of dynamical or stochastic processes on families of networks, but also includes community detection where a 'uniform' random graph model classifies nodes into communities depending on whether the graph under consideration deviates from typical realizations of the random graph model [36, 61].

Due to the aforementioned difficulty in choosing a particular random graph model to understand real world networks, we want to prove results regarding an assortment of random graph models in addition to the Uniform Model and focus on the related Chung-Lu model. Consequently in Chapter 4, we provide rigorous and novel consentration results regarding the relative error of the dominating eigenvalue of the adjacency matrix from the asymptotic limit in two cases, where either the likelihood of any two nodes sharing an edge vanishes in the limit or the quantity $\frac{\mathbf{a}\cdot\mathbf{b}}{S}$ diverges. The proof of the conjecture posed practical mathematical challenges as the vast majority of eigenvalue, eigenvector results deal with symmetric matrices; in contrast, we could not make such symmetry assumptions regarding our adjacency matrix. Furthermore, we emphasize that our results not only provide asymptotics for the dominating eigenvalue, but also provide useful concentration bounds for the dominating eigenvalue of the adjacency matrix in a graph with *finitely* many nodes. We then presented a novel approach for extending our asymptotic eigenvalue results to the Partitioned Chung-Lu random graph model where each partitioned submatrix has an expected degree sequence that follows the Chung-Lu random graph model, allowing for community structure in the network. Since the dominating eigenvalue of the adjacency matrix can correspond to the stability of solutions of certain dynamical processes, we demonstrated through numeric simulation the impact of partitioning or community structure regarding the onset of synchrony in the Kuramoto Model, a toy model for biological neural networks and the stability of the endemic state in an SIS (susceptible-infected-susceptible) model for epidemiological networks.

Our results addressing some of the challenges of modeling real world networks led to a number of open problems as well. Foremost in the uniform sampling problem, although our asymptotic results pertaining to the number of graphs that realize a given degree sequence can assist greatly in constructing samples from the uniform model, proving theoretical results regarding the actual computational complexity of implementing a particular sampling procedure remains an open problem. We also proved in this work that the uniform model converges to the Chung-Lu model. Since the convergence may not necessarily be fast, we want to stress the importance of choosing the correct random graph model along with the correct parameters when analyzing real world networks. Of course identifying the random graph model that appropriately emulates the empirically observed real world network seems like a highly non-trivial,partially ill-defined, and yet important problem for the practitioner. This challenge implicitly appears in selecting parameters for the Partitioned Chung-Lu Random Graph Model. When deciding the parameters for the Partitioned Chung-Lu random graph model to fit our empirically observed network, how should we partition our network and how many partitions should we choose?

Recall that we introduced the Partitioned Chung-Lu random graph model primarily because the model enabled us to extend our eigenvalue consentration results from the Chung-Lu model to allow for networks with possibly intricate community structure. Can we generalize our results pertaining to the concentration results of the dominating eigenvalue to other cases besides Chung-Lu type random graphs? Even though the Chung-Lu probabilites simplified the mathematical analysis regarding the asymptotic limit for the dominating eigenvalue, the main obstacle should stem from the fact that our adjacency matrix may not be symmetric, not the actual asymptotic limit of the dominating eigenvalue. Consequently, we expect that analogous eigenvalue concentration results should hold for other random graph models as well. We can also consider extensions of our eigenvalue results in terms of weighted directed graphs. In epidemiological networks certain individuals are more prone to infection than others, which we can model using weighted edges in our graph. Consequently, can we extend our eigenvalue concentration results to random graph models allowing for certain distributions of positive edge weights? We assert that these questions are exciting directions for further inquiry;but in any case, we anticipate that the results proven in this work: our easy to use sufficient conditions for guaranteeing graphicality of a collection of degree sequences, the novel procedure for constructing enumerative asymptotics of graphs with a prescribed degree sequence for sampling graphs from the uniform model, and the dominating eigenvalue concentration results for the adjacency matrix of directed graphs, will greatly assist us in our endeavor to better grasp the intricacies of modeling dynamical processes on real world networks.

BIBLIOGRAPHY

[1] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

[2] Noga Alon, Sonny Ben-Shimon, and Michael Krivelevich. A note on regular ramsey graphs. *Journal of Graph Theory*, 64(3):244–249, 2010.

[3] Richard P Anstee. Properties of a class of (0, 1)-matrices covering a given matrix. *Can. J. Math*, 34(2):438–453, 1982.

[4] Søren Asmussen and Peter W Glynn. *Stochastic simulation: algorithms and analysis*, volume 57. Springer Science & Business Media, 2007.

[5] Tiffany M Barnes and Carla D Savage. A recurrence for counting graphical partitions. *Electronic J. Comb*, 2:R11, 1995.

[6] Tiffany M Barnes and Carla D Savage. Efficient generation of graphical partitions. *Discrete applied mathematics*, 78(1):17–26, 1997.

[7] Alexander Barvinok. Matrices with prescribed row and column sums. *Linear Algebra and its Applications*, 436(4):820–844, 2012.

[8] Mohsen Bayati, Jeong Han Kim, and Amin Saberi. A sequential algorithm for generating random graphs. *Algorithmica*, 58(4):860–910, 2010.

[9] Edward A Bender and E Rodney Canfield. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296–307, 1978.

[10] Annabell Berger. A note on the characterization of digraph sequences. *arXiv preprint arXiv:1112.1215*, 2011.

[11] Annabell Berger. The connection between the number of realizations for degree sequences and majorization. *arXiv preprint arXiv:1212.5443*, 2012.

[12] Annabell Berger and Matthias Müller-Hannemann. Uniform sampling of digraphs with a fixed degree sequence. In *Graph theoretic concepts in computer science*, pages 220–231. Springer, 2010.

[13] Ivona Bezáková, Nayantara Bhatnagar, and Eric Vigoda. Sampling binary contingency tables with a greedy start. *Random Structures & Algorithms*, 30(1-2):168–205, 2007.

[14] Ivona Bezáková, Alistair Sinclair, Daniel Štefankovič, and Eric Vigoda. Negative examples for sequential importance sampling of binary contingency tables. *Algorithmica*, 64(4):606–620, 2012.

[15] Jose H Blanchet et al. Efficient importance sampling for binary contingency tables. *The Annals of Applied Probability*, 19(3):949–982, 2009.

[16] Joseph Blitzstein and Persi Diaconis. A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Mathematics*, 6(4):489–522, 2011.

[17] David Burstein, Franklin Kenter, Jeremy Kun, and Feng Shi. Interception in distance-vector routing networks. *To Appear in the Journal of Complex Networks*, 2016.

[18] David Burstein and Jonathan Rubin. Sufficient conditions for graphicality of bidegree sequences. *arXiv preprint arXiv:1511.02411*, 2015.

[19] Grant Cairns, Stacey Mendan, and Yuri Nikolayevsky. A sharp refinement of a result of alon, ben-shimon and krivelevich on bipartite graph vertex sequences. *arXiv preprint arXiv:1403.6307*, 2014.

[20] E Rodney Canfield, Catherine Greenhill, and Brendan D McKay. Asymptotic enumeration of dense 0–1 matrices with specified line sums. *Journal of Combinatorial Theory, Series A*, 115(1):32–66, 2008.

[21] Kamalika Chaudhuri, Fan Chung, and Alexander Tsiatas. Spectral clustering of graphs with general degrees in the extended planted partition model. *Journal of Machine Learning Research*, 2012:1–23, 2012.

[22] Sanjeev Chauhan, Michelle Girvan, and Edward Ott. Spectral properties of networks with community structure. *Physical Review E*, 80(5):056114, 2009.

[23] Ningyuan Chen, Mariana Olvera-Cravioto, et al. Directed random graphs with given degree distributions. *Stochastic Systems*, 3(1):147–186, 2013.

[24] Wai-Kai Chen. On the realization of a (p, s)-digraph with prescribed degrees. *Journal of the Franklin Institute*, 281(5):406–422, 1966.

[25] Yuguo Chen, Persi Diaconis, Susan P Holmes, and Jun S Liu. Sequential monte carlo methods for statistical analysis of tables. *Journal of the American Statistical Association*, 100(469):109–120, 2005.

[26] Leonid Chindelevitch, Po-Ru Loh, Ahmed Enayetallah, Bonnie Berger, and Daniel Ziemek. Assessing statistical significance in causal graphs. *BMC bioinformatics*, 13(1):35, 2012.

[27] Fan Chung and Linyuan Lu. Connected components in random graphs with given expected degree sequences. *Annals of combinatorics*, 6(2):125–145, 2002.

[28] Fan Chung and Linyuan Lu. The average distance in a random graph with given expected degrees. *Internet Mathematics*, 1(1):91–113, 2004.

[29] Fan Chung, Linyuan Lu, and Van Vu. Eigenvalues of random power law graphs. *Annals of Combinatorics*, 7(1):21–33, 2003.

[30] Fan Chung and Mary Radcliffe. On the spectra of general random graphs. *the electronic journal of combinatorics*, 18(1):P215, 2011.

[31] Geir Dahl and Truls Flatberg. A remark concerning graphical sequences. *Discrete mathematics*, 304(1):62–64, 2005.

[32] Charo I Del Genio, Hyunju Kim, Zoltan Toroczkai, and Kevin E Bassler. Efficient and exact sampling of simple graphs with given arbitrary degree sequence. *PloS one*, 5(4):e10012, 2010.

[33] Paul Erdős and Lawrence Bruce Richmond. On graphical partitions. *Combinatorica*, 13(1):57–63, 1993.

[34] Péter L Erdos, István Miklós, and Zoltán Toroczkai. A simple havel-hakimi type algorithm to realize graphical degree sequences of directed graphs. *Electronic Journal of Combinatorics*, 17(1):R66, 2010.

[35] Silvio C Ferreira, Claudio Castellano, and Romualdo Pastor-Satorras. Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results. *Physical Review E*, 86(4):041125, 2012.

[36] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3):75–174, 2010.

[37] Delbert Ray Fulkerson et al. Zero-one matrices with zero trace. *Pacific J. Math*, 10(3):831–836, 1960.

[38] David Gale et al. A theorem on flows in networks. *Pacific J. Math*, 7(2):1073–1082, 1957.

[39] Ayalvadi Ganesh, Laurent Massoulié, and Don Towsley. The effect of network topology on the spread of epidemics. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, volume 2, pages 1455–1466. IEEE, 2005.

[40] Jennifer L Gardy, James C Johnston, Shannan J Ho Sui, Victoria J Cook, Lena Shah, Elizabeth Brodkin, Shirley Rempel, Richard Moore, Yongjun Zhao, Robert Holt, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *New England Journal of Medicine*, 364(8):730–739, 2011.

[41] Aristides Gionis, Heikki Mannila, Taneli Mielikäinen, and Panayiotis Tsaparas. Assessing data mining results via swap randomization. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(3):14, 2007.

[42] Catherine Greenhill, Brendan D McKay, and Xiaoji Wang. Asymptotic enumeration of sparse 0–1 matrices with irregular row and column sums. *Journal of Combinatorial Theory, Series A*, 113(2):291–324, 2006.

[43] S Louis Hakimi. On the degrees of the vertices of a directed graph. *Journal of the Franklin Institute*, 279(4):290–308, 1965.

[44] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[45] S Itzkovitz, R Milo, N Kashtan, MEJ Newman, and U Alon. Reply to comment on subgraphs in random networks. *Physical Review E*, 70(5):058102, 2004.

[46] Shalev Itzkovitz, Ron Milo, Nadav Kashtan, Guy Ziv, and Uri Alon. Subgraphs in random networks. *Physical review E*, 68(2):026127, 2003.

[47] Svante Janson. Poisson approximation for large deviations. *Random Structures & Algorithms*, 1(2):221–229, 1990.

[48] Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.

[49] Hyunju Kim, Charo I Del Genio, Kevin E Bassler, and Zoltán Toroczkai. Constructing and sampling directed graphs with given degree sequences. *New Journal of Physics*, 14(2):023012, 2012.

[50] Daniel J Kleitman and Da-Lun Wang. Algorithms for constructing graphs and digraphs with given valences and factors. *Discrete Mathematics*, 6(1):79–88, 1973.

[51] Can M Le and Roman Vershynin. Concentration and regularization of random graphs. *arXiv preprint arXiv:1506.00669*, 2015.

[52] Linyuan Lu and Xing Peng. Spectra of edge-independent random graphs. *arXiv preprint arXiv:1204.6207*, 2012.

[53] Xuesong Lu and Stéphane Bressan. Generating random graphic sequences. In *Database Systems for Advanced Applications*, pages 570–579. Springer, 2011.

[54] Brendan D McKay. Asymptotics for 0-1 matrices with prescribed line sums. *Enumeration and Design,(Academic Press, 1984)*, pages 225–238, 1984.

[55] Brendan D McKay et al. Subgraphs of random graphs with specified degrees. In *Proceedings of the International Congress of Mathematicians*, volume 4, pages 2489–2501, 2010.

[56] Brendan D McKay and Nicholas C Wormald. Asymptotic enumeration by degree sequence of graphs with degreeso (n 1/2). *Combinatorica*, 11(4):369–382, 1991.

[57] Jeffrey W Miller. Reduced criteria for degree sequences. *Discrete Mathematics*, 313(4):550–562, 2013.

[58] Jeffrey W Miller, Matthew T Harrison, et al. Exact sampling and counting for fixed-margin matrices. *The Annals of Statistics*, 41(3):1569–1592, 2013.

[59] Ron Milo, Nadav Kashtan, Shalev Itzkovitz, Mark EJ Newman, and Uri Alon. On the uniform generation of random graphs with prescribed degree sequences. *arXiv preprint cond-mat/0312028*, 2003.

[60] Raj Rao Nadakuditi and Mark EJ Newman. Spectra of random graphs with arbitrary expected degrees. *Physical Review E*, 87(1):012803, 2013.

[61] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582, 2006.

[62] Jennifer M Nolan, Vijay Sivaraman, Carla D Savage, and Pranav K Tiwari. Graphical basis partitions. *Graphs and Combinatorics*, 14(3):241–261, 1998.

[63] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, 2015.

[64] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics in finite size scale-free networks. *Physical Review E*, 65(3):035108, 2002.

[65] Tiago P Peixoto. Eigenvalue spectra of modular networks. *Physical review letters*, 111(9):098701, 2013.

[66] Boris Pittel. Confirming two conjectures about the integer partitions. *Journal of Combinatorial Theory, Series A*, 88(1):123–135, 1999.

[67] Andrew Pomerance, Edward Ott, Michelle Girvan, and Wolfgang Losert. The effect of network topology on the stability of discrete state models of genetic control. *Proceedings of the National Academy of Sciences*, 106(20):8209–8214, 2009.

[68] Mason A Porter and James P Gleeson. Dynamical systems on networks: a tutorial. *arXiv preprint arXiv:1403.7663*, 2014.

[69] A Ramachandra Rao, Rabindranath Jana, and Suraj Bandyopadhyay. A markov chain monte carlo method for generating random (0, 1)-matrices with given marginals. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 225–242, 1996.

[70] Ronald C Read. *Some Enumeration Problems in Gaph Theory*. 1958.

[71] Juan G Restrepo, Edward Ott, and Brian R Hunt. Onset of synchronization in large networks of coupled oscillators. *Physical Review E*, 71(3), 2005.

[72] Juan G Restrepo, Edward Ott, and Brian R Hunt. Emergence of synchronization in complex networks of interacting dynamical systems. *Physica D: Nonlinear Phenomena*, 224(1):114–122, 2006.

[73] Juan G Restrepo, Edward Ott, and Brian R Hunt. Approximating the largest eigenvalue of network adjacency matrices. *Physical Review E*, 76(5):056119, 2007.

[74] +Peter+D. Roopnarine and +Kenneth+D." Angielczyk. Community stability and selective extinction during the permian-triassic mass extinction. *Science*, 350(6256):90–93, 2015.

[75] HJ Ryser. Combinatorial properties of matrices of zeros and ones. In *Classic Papers in Combinatorics*, pages 269–275. Springer, 2009.

[76] Wolfgang E Schlauch, Katharina A Zweig, et al. Influence of the null-model on motif detection. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pages 514–519. ACM, 2015.

[77] Tiziano Squartini and Diego Garlaschelli. Analytical maximum-likelihood method to detect patterns in real networks. *New Journal of Physics*, 13(8):083001, 2011.

[78] Norman D Verhelst. An efficient mcmc algorithm to sample binary matrices with fixed marginals. *Psychometrika*, 73(4):705–728, 2008.

[79] Xiao Zhang, Raj Rao Nadakuditi, and Mark EJ Newman. Spectra of random graphs with community structure and arbitrary degrees. *Physical Review E*, 89(4):042816, 2014.

[80] Liqiong Zhao, Bryce Beverlin, Theoden Netoff, and Duane Q Nykamp. Synchronization from second order network connectivity statistics. *Frontiers in computational neuroscience*, 5(28):1–16, 2011.

[81] Yunpeng Zhao, Elizaveta Levina, and Ji Zhu. Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, pages 2266–2292, 2012.

[82] Igor E Zverovich and Vadim E Zverovich. Contributions to the theory of graphic sequences. *Discrete Mathematics*, 105(1):293–303, 1992.