Degree Sequences of Random Bipartite Graphs

Fiona Skerman

2010

A thesis submitted as a partial requirement of a Degree of

Bachelor of Philosophy with Honours.

Department of Mathematics

The Australian National University

October 28, 2010

Abstract

This concerns three random bipartite graph models. For each random graph model a binomially based model is explicitly constructed which has the property that for 'most' degree sequences $(\boldsymbol{s}, \boldsymbol{t})$, the probability of $(\boldsymbol{s}, \boldsymbol{t})$ in the graph model is asymptotically approximated by the probability of $(\boldsymbol{s}, \boldsymbol{t})$ in the binomial model. This allows us to prove Theorems 7.4 and 7.5 which are the bipartite analogue of Theorem 2.6 in the paper by McKay and Wormald [MW97]. This construction is new as are Theorems 7.4 and 7.5.

Acknowledgements

It is an honor for me to thank both my co-supervisors Prof. Brendan Mckay and Dr. Judy-anne Osborn. I would like to thank Brendan Mckay for his unending fountain of mathematical insight and interesting ideas. I would also like to show my gratitude to Judy-anne Osborn for her time and help in writing the thesis. Their patience and understanding made this thesis possible.

I am also indebted to Dr. Adam Rennie for his advice on the art of mathematical writing and thank him for his help. Also to my fellow honours students, especially Brendan Fong, Howard Chuang, Tom Phelan, Ross Atkins, Wilson Ong and David Shellard with whom I shared an office.

Additionally, I would like acknowledge my Mother and Father who have helped and inspired me in ways too numerous to mention. Thanks also goes to the university for their generous support of my studies via the National Undergraduate Scholarship.

Contents

Introduction

This thesis concerns random bipartite graphs.

We define the key parameters of bipartite graphs. This follows the notation of [GM09].

For $m, n \in \mathbb{N}$, define $N = mn$ and $I_{m,n} = \{0, 1, \ldots, n\}^m \times \{0, 1, \ldots, m\}^n$. We define an $(m+n)$-tuple $(\boldsymbol{s}, \boldsymbol{t}) \in I_{m,n}$ and denote the components of the m-tuple $\boldsymbol{s}$, by $s_1, s_2, \ldots, s_m$ and the n-tuple $\boldsymbol{t}$, by $t_1, \ldots, t_n$. Define the averages $s = \frac{1}{n} \sum_{j=1}^{m} s_j$ and $t = \frac{1}{m} \sum_{k=1}^{n} t_k$.

We can now define a bipartite graph on (m, n) vertices. It has a left vertex set $U = \{u_1, u_2, \ldots, u_m\}$ (drawn white) and right vertex set $V = \{v_1, v_2, \ldots, v_n\}$ (drawn black). We also define an edge set $E \subseteq \{\ (u_j, v_k) \ : \ u_j \in U,\ v_k \in V\}$. We call the number of edges incident with a vertex the *degree* of that vertex. Denote the degree of u_j by s_j and the degree of v_k by t_k. The *degree sequence* of the graph is then $(\boldsymbol{s}, \boldsymbol{t})$ as defined above.

We also define the *edge-density*, λ, of a bipartite graph. Observe that the number of edges in a bipartite graph can be determined by counting up the degrees of either side, so $\#\text{edges} = \sum_j s_j =: \lambda mn$.

We illustrate these concepts in Figure 1.

There is a natural bijection between bipartite graphs on (m, n) vertices and $m \times n$ binary matrices where an edge between *white* vertex u_j and *black* vertex v_k corresponds to a '1' in the j^{th} row and k^{th} edge. Likewise the absence of an edge indicates a '0' in the corresponding position of the binary matrix. Hence this work can be thought of in this context. For instance, the degree of vertex u_j a graph would become the row sum of the j^{th} row in the corresponding matrix.

This thesis concerns probability spaces over bipartite graphs. Let m, n be integers and let $p = p(m, n)$ be a function into $[0, 1]$. We define a random bipartite graph with m *white* vertices and n *black* vertices of the other. In this random graph, each of the mn possible edges is present independently with probability $p(m, n)$. We call the resulting probability space the graph p-model and denote it by $\mathcal{G}_p$.

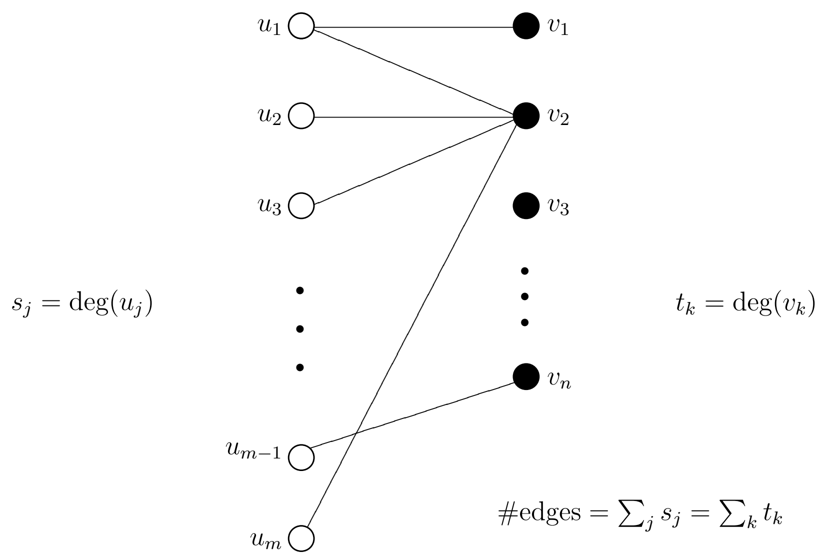


Figure 1: This gives a representation of a possible bipartite graph on (m, n) vertices. In the graph there are four edges incident with *black* vertex v_2 and so $t_2 = 4$.

The graph p-model, $\mathcal{G}_p$ defined above is one way to define a random bipartite graph. We also define two other models for random graphs, the graph *edge*-model, denoted $\mathcal{G}_M$, and the graph *half*-model, denoted $\mathcal{G}_{\boldsymbol{t}}$. In the graph *edge*-model, let M be an integer between zero and mn. We then set each possible bipartite graph $G(m, n)$ with M edges to have equal probability.

The graph *half*-model gets its name because we specify the degrees of the n *black* vertices. All bipartite graphs whose *black* degrees match those prescribed are then chosen with equal probability. We thus have three different random graph models, $\mathcal{G}_p, \mathcal{G}_M$ and $\mathcal{G}_{\boldsymbol{t}}$ i.e. three different ways of choosing a graph at random.

Of interest is the probability distribution of the degree sequence of a random graph as the number of vertices becomes very large. We consider the case when the ratio of *white* to *black* vertices is close to one. This is studied for each random graph model.

Hence our problem becomes to find the probability of a graph having a given degree sequence $(\boldsymbol{s}, \boldsymbol{t})$ in each of our graph models. This we could calculate if we knew the number of bipartite graphs that can be constructed with degree sequence $(\boldsymbol{s}, \boldsymbol{t})$. Our starting point is a theorem by Greenhill and McKay [GM09] which has done this enumeration for bipartite graphs that satisfy certain conditions on the degree sequence. We term such degree sequences (ε, a)-*regular*.

We display the enumeration formula in the following theorem. Even without understanding every term seeing the general form will help explain the direction of this thesis. We refer the reader ahead to Theorem 2.8 for a full statement of the result. (Below, $|B(\boldsymbol{s},\boldsymbol{t})|$ indicates the number of bipartite graphs with degree sequence $(\boldsymbol{s},\boldsymbol{t})$ and the constant b in the error term is in the interval $(0,\frac{1}{2})$.)

Theorem 0.1 (Greenhill and McKay)**.** *For* (ε,a)-regular* *degree sequences:*

$$|B(\boldsymbol{s},\boldsymbol{t})| = \binom{mn}{\lambda mn}^{-1} \prod_{j=1}^{m}\binom{n}{s_j}\prod_{k=1}^{n}\binom{m}{t_k} \\ \times \exp\left(-\tfrac{1}{2}\Big(1-\frac{\sum_j (s_j-s)^2}{\lambda(1-\lambda)mn}\Big)\Big(1-\frac{\sum_k (t_k-t)^2}{\lambda(1-\lambda)mn}\Big) + O(n^{-b})\right) \tag{0.1}$$

A natural question which arises given this counting result is how common these well-behaved (ε,a)*-regular* degree sequences are in our random graph models. We show that (ε,a)*-regular* degree sequences account for the bulk of the probability space in each of the three random graph models.

Result 1[†]: (ε,a)-regular degree Sequences

For each random graph model, asymptotically,
almost all degree sequences are (ε,a)-regular.

Note that if a degree sequence $(\boldsymbol{s},\boldsymbol{t})$ is both (ε,a)*-regular* and

$$\Big(1-\frac{\sum_j (s_j-s)^2}{\lambda(1-\lambda)mn}\Big)\Big(1-\frac{\sum_k (t_k-t)^2}{\lambda(1-\lambda)mn}\Big) = O(n^{-b}),$$

then this simplifies the formula by McKay and Greenhill above. We call such sequences non-*pathological*.

Result 2[‡]: *Pathological* degree sequences

For each random graph model, asymptotically, the probability
a degree sequence is pathological *tends to zero.*

After proving the previous two results we are in a good position. In each model, except for rare (i.e. *pathological*) degree sequences we have an asymptotic count for the number of

[*] (ε,a)-regular degree sequences, see Definition 2.7

[†] For the precise statement of this result for graph models $\mathcal{G}_p$, $\mathcal{G}_M$ and $\mathcal{G}_{\boldsymbol{t}}$ see Lemmas 4.8, 5.1 and 6.1 respectively.

[‡] We show this result for models $\mathcal{G}_p$, $\mathcal{G}_M$ and $\mathcal{G}_{\boldsymbol{t}}$ in Theorems 4.9, 5.2 and 6.20 respectively.

bipartite graphs with that degree sequence. From these counts we derive the approximate probability of finding a graph with that given non-*pathological* degree sequence. To get to this point comprises the bulk of the thesis, up until the end of Chapter 6.

For all but *pathological* degree sequence the formula in (0.1) simplifies and these *pathological* degree sequences are rare in our random bipartite graph models. This means that for 'most' degree sequences we are able to find a simple asymptotic count for the number of bipartite graphs with that given degree sequence.

We explicitly construct binomial models $\mathcal{B}_M, \mathcal{B}_{\boldsymbol{t}}$ and $\mathcal{V}_p$ based on binomially distributed random variables subject to certain constraints[§]. Let the binomial models $\mathcal{B}_M, \mathcal{B}_{\boldsymbol{t}}$ and $\mathcal{V}_p$ correspond to the graph models $\mathcal{G}_M, \mathcal{G}_{\boldsymbol{t}}$ and $\mathcal{G}_p$ respectively.

Then these binomial models have the property that, for each degree sequence, the probability of that degree sequence occurring in the random graph model and the probability of it occurring in the corresponding binomial model are very close. This is an original result and is the culmination of all calculations in the thesis.

We summarise this result below. Each random graph model can be approximated by one of these newly defined binomial models in the following fashion.

Result 3[¶]

*There exist probability spaces, $\mathcal{B}_M, \mathcal{B}_{\boldsymbol{t}}$ and $\mathcal{V}_p$ such that for any non-*pathological *degree sequence $(\boldsymbol{s}, \boldsymbol{t})$,*

$$\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{s}, \boldsymbol{t}) = \mathbb{P}_{\mathcal{G}_M}(\boldsymbol{s}, \boldsymbol{t})(1 + O(n^{-b}))$$

$$\mathbb{P}_{\mathcal{B}_{\boldsymbol{t}}}(\boldsymbol{s}, \boldsymbol{t}) = \mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(\boldsymbol{s}, \boldsymbol{t})(1 + O(n^{-b}))$$

$$\mathbb{P}_{\mathcal{V}_p}(\boldsymbol{s}, \boldsymbol{t}) = \mathbb{P}_{\mathcal{G}_p}(\boldsymbol{s}, \boldsymbol{t})(1 + O(n^{-b})).$$

(An explicit construction is given for each probability space.)

Each of these newly constructed binomial spaces is based on binomial random variables subject to constraints on their sum. This has interesting theoretical implications. Consider the degree of a fixed vertex in the random bipartite graph p-model, $\mathcal{G}_p$. This vertex alone has a binomial distribution. Moreover, vertices of the same colour have independent degrees. However, vertices of different colours have degrees which are dependent. Our results will enable us to begin to quantify the magnitude of this dependence.

[§] For the constructions see Definition 7.3 of $\mathcal{B}_M$, Definition 7.5 of $\mathcal{B}_{\boldsymbol{t}}$ and Definition 7.7 of $\mathcal{V}_p$.

[¶] This result appears as three separate Theorems in this thesis. We prove $\mathcal{B}_M \sim \mathcal{G}_M$ in Theorem 7.1, $\mathcal{B}_{\boldsymbol{t}} \sim \mathcal{G}_{\boldsymbol{t}}$ in Theorem 7.2 and lastly $\mathcal{V}_p \sim \mathcal{G}_p$ in Theorem 7.10.

Part I

Literature review and background theory

Chapter 1

Preliminaries

1.1 Definitions

1.1.1 Probability

To be able to define what is meant by a random graph and also to make a meaningful survey of previous results we need some probability theory. We give the necessary probability theory background for the content of the preliminary section (Part I). Later in Chapter 3 we detail the remaining probability results that are needed for the main body of the thesis.

Many objects and functions studied in probability are special cases of familiar concepts from measure theory. We set out the some of the definitions we will require.

Definition 1.1 (Probability space)**.** *Suppose we have a measure space which is defined by the triple* (Ω, Σ, μ)*. If* $\mu(\Omega) = 1$ *then the triple* (Ω, Σ, μ) *is a* probability space *and* μ *is a* probability measure*.*

Observe that if the measure μ satisfies $\mu(\Omega) = k < \infty$, then it naturally determines a probability measure $\mu'(X) := \frac{1}{k}\mu(X)$. This is referred to as *normalisation*.

In a probability space we will often denote the measure by $\mathbb{P}$.

Definition 1.2 (Random variable)**.** *Let* $(\Omega, \Sigma, \mathbb{P})$ *be a probability space. Suppose we have a function* $X : \Omega \to \mathbb{R}$*, then* X *is a* random variable *if for each* $r \in \mathbb{R}$*,*

$$\{\omega : X(w) \leq r\} \in \Sigma \tag{1.1}$$

In this thesis we will deal only probability spaces $(\Omega, \Sigma, \mathbb{P})$ where $|\Sigma|$ is finite and $\Sigma = \mathcal{P}(\Omega)$. In this special case every subset of Ω is an element of Σ and so the requirement (1.1) will hold for any function $X : \Omega \to \mathbb{R}$.

The capital letters X, Y will be used to denote random variables, whilst the lowercase x, y used to denote particular events in Σ. We denote the probability of the random variable X taking value x as $\mathbb{P}(X = x)$.

Definition 1.3 (Expected value)**.** *Let X, be a* random variable *defined on the finite probability space* $(\Omega, \Sigma, \mathbb{P})$*. Then the* expected value *of X is defined by*

$$\mathbb{E}(X) := \sum_{x \in \Omega} x \mathbb{P}(X = x).$$

In this thesis we will find that some of the random variables defined on our random graph models have probability distributions that are well-known. For example, let S_j be the random variable which returns the degree of the j^{th} *white* vertex in $\mathcal{G}_p$. For any $1 \leq j \leq m$ we note in (2.2) that the degree of *white* vertex u_j has a binomial distribution with parameters (n, p), in the graph p-model, $\mathcal{G}_p$.

We define the probability distributions we will encounter. In the following definitions we let $[n] = \{0, 1, \ldots, n\}$.

Definition 1.4 (Binomial distribution)**.** *Let* $([n], \mathcal{P}([n]), \mathbb{P})$ *be a probability space and X be a real-valued random variable on this space. Then we say X is* binomially distributed with parameters (n, p)*, if for each* $0 \leq r \leq n$*,*

$$\mathbb{P}(X = r) = \binom{n}{r} p^r (1 - p)^{n-r}.$$

Definition 1.5 (Poisson distribution)**.** *Let* $(\Omega, \Sigma, \mathbb{P})$ *be a probability space and X be a real-valued random variable on this space. Then we say X has a* Poisson distribution with parameter c*, if for each* $r \in \mathbb{N}$*,*

$$\mathbb{P}(X = r) = \frac{1}{r!} e^{-c} c^r.$$

1.1.2 Random bipartite graphs models

In this section we define our models and state some of their elementary properties. First we will need a little more notation than has already been defined on p.v. We will then define the random bipartite graph models, i.e. probability spaces over bipartite graphs.

Notation The total number of edges in a bipartite graph can be determined from its degree sequence by summing the degrees of either side. Hence a necessary condition for an $(m+n)$-tuple, $(\boldsymbol{s}, \boldsymbol{t})$, to correspond to a degree sequence of a bipartite graph is that $\sum_j s_j = \sum_k t_k$. With this motivation we define the following subsets of $I_{m,n}$,

$$E_{m,n} := \{(\boldsymbol{s}, \boldsymbol{t}) \in I_{m,n} \mid \sum_j s_j = \sum_k t_k\} \tag{1.2}$$

$$E_{m,n,M} := \{(\boldsymbol{s}, \boldsymbol{t}) \in E_{m,n} \mid \sum_j s_j = M\} \qquad \text{(for } 0 \leq M \leq mn\text{)} \tag{1.3}$$

$$E_{m,n\boldsymbol{t}'} := \{(\boldsymbol{s}, \boldsymbol{t}) \in E_{m,n} \mid \boldsymbol{t} = \boldsymbol{t}'\} \qquad \text{(for } n\text{-tuples } \boldsymbol{t}'\text{)}. \tag{1.4}$$

Random variables This thesis will consider the following random variables on graphs.

For $1 \leq j \leq m$, let S_j be the random variable which returns the degree of the white *vertex u_j and define $S := \frac{1}{m}\sum_j S_j$.*

For $1 \leq k \leq n$, let T_k be the random variable which returns the degree of the black *vertex v_k and define $T := \frac{1}{n}\sum_k T_k$.*

Define $\boldsymbol{S} := (S_1, \ldots, S_m)$ and $\boldsymbol{T} := (T_1, \ldots, T_n)$.

The models The notation we defined for our random bipartite graph models is consistent with that used for the general random graph models considered in [MW97]. There are two main random models for general graphs[1], $\hat{\mathcal{G}}_p$ and $\hat{\mathcal{G}}_M$. Refer to Section 2.1 for a definition of these models and some important results in general (non-bipartite) random graph theory.

As direct analogues of these in the bipartite case we define bipartite random graph models; the graph p-model, $\mathcal{G}_p$, and the graph *edge*-model, $\mathcal{G}_M$. We also define a model unique to bipartite graphs, the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$.

These three models are all probability spaces with the same domain, the set of all bipartite graphs on (m, n) vertices. Denote this domain by $\mathcal{B}_{m,n}$ and write $2^{\mathcal{B}_{m,n}}$ for the power set of $\mathcal{B}_{m,n}$. Then we can express the probability spaces formally as $(\mathcal{B}_{m,n}, 2^{\mathcal{B}_{m,n}}, \phi)$, where only the probability measure, ϕ, differs between the three models.

Graph p-model, $\mathcal{G}_p$ This model of a random bipartite graph appears in [GLS99], [Pal88] and [Pal84].

[1] Refer to the first section, *The Basic Models*, in the chapter on Random Graphs in the book, *Modern Graph Theory*, by Bollobás [Bol98]

Definition 1.6 (Graph p-model, $\mathcal{G}_p$)**.** The graph p-model, $\mathcal{G}_p(m,n) = (\mathcal{B}_{m,n}, 2^{\mathcal{B}_{m,n}}, \mathbb{P}_{\mathcal{G}_p})$ has domain the set of all labelled bipartite graphs on (m,n) vertices . A graph is chosen at random by selecting each of the possible mn edges independently with probability p.

Hence in this model, the probability of a particular graph H with (m,n) vertices and $|H|$ edges is, $\mathbb{P}_{\mathcal{G}_p}(H) = p^{|H|}q^{mn-|H|}$.

For any degree sequence $(\boldsymbol{s},\boldsymbol{t}) \in E_{m,n}$ we are interested in the probability that a random graph in the graph p-model, $\mathcal{G}_p$, has this degree sequence. The *probability of the degree sequence* $(\boldsymbol{s},\boldsymbol{t})$ *in the graph p-model* written,

$$\mathbb{P}_{\mathcal{G}_p}\big(\boldsymbol{s},\boldsymbol{t}\big) := \mathbb{P}_{\mathcal{G}_p}\big(H : \textit{the degree sequence of } H \textit{ is } (\boldsymbol{s},\boldsymbol{t})\big).$$

If we represent a bipartite graph as a binary matrix then the $\mathcal{G}_p(m,n)$ model corresponds to an $m \times n$ matrix where each entry is independently chosen to be a 1 with probability p (and a zero with probability $q = 1-p$). Thus for an $(m+n)$-tuple $(\boldsymbol{s},\boldsymbol{t})$, $\mathbb{P}_{\mathcal{G}_p}(\boldsymbol{s},\boldsymbol{t})$ is the probability that a binary matrix chosen at random in this way will have row sums $\boldsymbol{s}$ and column sums $\boldsymbol{t}$.

Graph *edge*-model, $\mathcal{G}_M$ This model of a random bipartite graph appears in [GLS99], [ER61] and [Pal84].

Definition 1.7 (Graph *edge*-model, $\mathcal{G}_M$)**.** The graph *edge*-model$\mathcal{G}_M = \mathcal{G}_M(m,n,M)$ has support the set of all labelled bipartite graphs on (m,n) vertices with M edges. We define each of the $\binom{mn}{M}$ different graphs in the support to be equiprobable.

Hence in the graph *edge*-model, the probability of a particular graph H with M edges vertices is $\mathbb{P}_{\mathcal{G}_M}(G) = \binom{mn}{M}^{-1}$. Also notice that choosing a graph at random in the graph *edge*-model, $\mathcal{G}_M$ corresponds to placing M ones in an $m \times n$ binary matrix where each of the $\binom{mn}{M}$ arrangements are equally likely.

For any degree sequence $(\boldsymbol{s},\boldsymbol{t}) \in E_{m,n,M}$ we are interested in the probability that our random graph in the graph *edge*-model, $\mathcal{G}_M$, has this degree sequence. For the *probability of the degree sequence* $(\boldsymbol{s},\boldsymbol{t})$ *in the graph* edge-*model* we will write,

$$\mathbb{P}_{\mathcal{G}_M}\big(\boldsymbol{s},\boldsymbol{t}\big) := \mathbb{P}_{\mathcal{G}_M}(\{H : \textit{the degree sequence of } H \textit{ is } (\boldsymbol{s},\boldsymbol{t})\}).$$

A consequence of the definitions of $\mathcal{G}_p$ and $\mathcal{G}_M$ is that for $0 \le M \le mn$ and $0 < p < 1$ then given any event $A \subset 2^{\mathcal{B}_{m,n}}$ we have

$$\mathbb{P}_{\mathcal{G}_M}(A) = \mathbb{P}_{\mathcal{G}_p}\big(A \mid M \text{ edges}\big).$$

We now give an example. In Figure 1.1 we show the three bipartite graphs with degree sequence $((1,3,1,1),(2,1,3))$. We subsequently calculate the probability of this degree sequence occurring in each of the two random graph models we have defined.

Bipartite graphs with vertices $(m,n) = (4,3)$

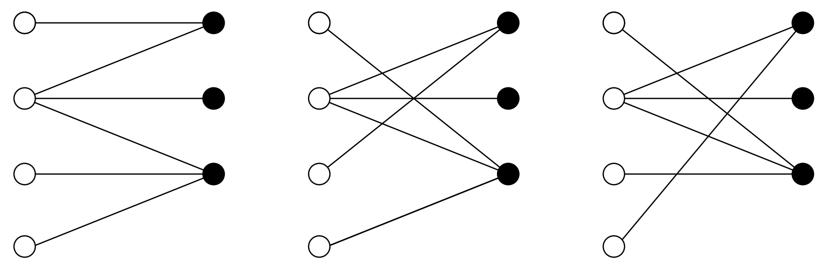


Figure 1.1: The three possible bipartite graphs with degree sequence ((1,3,1,1), (2,1,3)).

The total number of bipartite graphs on $(4,3)$ is $2^{12} = 4096$, of these, $\binom{12}{6} = 924$ of these have precisely 6 edges. Hence,

$$\mathbb{P}_{\mathcal{G}_{p=\frac{1}{2}}}((1,3,1,1),(2,1,3)) = \frac{3}{4096}$$

$$\mathbb{P}_{\mathcal{G}_{M=6}}((1,3,1,1),(2,1,3)) = \frac{3}{924} = \frac{1}{308}$$

Graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$ Unique to bipartite graphs we can define a third random graph model. In the graph *half*-model we fix the degrees of the vertices on one side. Our convention will be to fix the degrees of the *black* vertices.

Definition 1.8 (Graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$.)**.** The **graph *edge*-model** $\mathcal{G}_{\boldsymbol{t}} = \mathcal{G}_{\boldsymbol{t}}(m,n,\boldsymbol{t})$ has support the set of all labelled bipartite graphs on (m,n) vertices such that the degrees of the *black* vertices agree with those in a specified n-tuple $\boldsymbol{t}$. We define each of the different graphs in the support to be equiprobable.

The graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$, is equivalent to the probability space over all $m \times n$ binary matrices with column sums $\boldsymbol{t}$ where each matrix is weighted equally.

Observe that in the graph *half*-model, the probability of a particular graph H on (m,n) vertices with *black* degree sequence matching the n-tuple $\boldsymbol{t} = (t_1,\ldots,t_n)$ is $\left(\binom{m}{t_1}\binom{m}{t_2}\ldots\binom{m}{t_n}\right)^{-1}$.

Let $M = \sum_k t_k$. For any degree sequence $(\boldsymbol{s}, \boldsymbol{t}) \in E_{m,n,M}$ we are interested in the probability that our random graph in the *half*-model, $\mathcal{G}_{\boldsymbol{t}}$, has this degree sequence. The *probability of the degree sequence* $(\boldsymbol{s}, \boldsymbol{t})$ *in the graph* half-*model* is written,

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(\boldsymbol{s}, \boldsymbol{t}) := \mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(\{H : \textit{the degree sequence of } H \textit{ is } (\boldsymbol{s}, \boldsymbol{t})\}).$$

A consequence of this definition is that for any n-tuple $\boldsymbol{t}$ and $0 < p < 1$ then given any $A \subset I_{m,n,\boldsymbol{t}}$ we have

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(A) = \mathbb{P}_{\mathcal{G}_p}\big(A \mid \boldsymbol{T} = \boldsymbol{t}\big).$$

Example Suppose we want to find the probability of degree sequence $((1,3,1,1),(2,1,3))$ given we know that the degree sequence of the *black* vertices is $(2,1,3)$, i.e. we want to calculate $\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}=(2,1,3)}}((1,3,1,1),(2,1,3))$.

We first calculate the total number of different bipartite graphs on $(4,3)$ vertices with the prescribed degrees $(2,1,3)$ for the *black* vertices. Consider vertex v_1, there are two edges incident with this vertex creating $\binom{4}{2} = 6$ distinguishable positions for these two edges. Hence there are $\binom{4}{2}\binom{4}{1}\binom{4}{3} = 96$ distinguishable bipartite graphs with the required *black* vertices degree sequence. As illustrated by figure 1.1 on p6 there are three graphs with *white* vertex degree sequence $(1,3,1,1)$ and *black* vertex degree sequence $(2,1,3)$. Hence,

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}=(2,1,3)}}((1,3,1,1),(2,1,3)) = \frac{3}{96} = \frac{1}{32}.$$

Other models An interesting and general model is suggested by Palka in [Pal87]. A random graph is chosen with reference to an *initial graph* in the following way. Each edge of the *initial graph* remains with probability p (i.e. is deleted with probability $q = 1 - p$) and no new edges are added to the graph.

There are two important special cases. In the bipartite case, if we take the *initial graph*, to be the complete graph on n vertices then this is the $\hat{\mathcal{G}}_p(n)$ model defined earlier and if we take the *initial graph* to be the complete bipartite graph on (m, n) vertices then we have the bipartite graph p-model, $\mathcal{G}_p$, which we introduced in Definition 1.6.

1.1.3 Big 'O' notation

We will be dealing with the asymptotic properties of real functions of two variables m and n.

Definition 1.9 ($f(m,n) = O(g(m,n))$ as $m,n \to \infty$ subject to C)**.** *Suppose that $f, g : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ and C is a predicate on $\mathbb{N} \times \mathbb{N}$. Then we say '$f(m,n) = O(g(m,n))$ as $m,n \to \infty$ subject to C' if*
if there are constants $A, N > 0$ such that

$$|f(m,n)| \leq A|g(m,n)| \text{ whenever } m,n \geq N \text{ and } (m,n) \text{ satisfies } C.$$

The corresponding term for little 'o', $f(m,n) = o(g(m,n))$ as $m,n \to \infty$ subject to C is defined analogously.

In our case m and n will be the number of *white* and *black* vertices respectively. We will consider asymptotic results, as our graphs get large, where we also require that the two sets of vertices are not too different in size.

The following section is not required for understanding. It gives possible motivation for why one might want to define random bipartite graphs in the form of a toy example. A detailed calculation in the $\mathcal{G}_{\boldsymbol{t}}$ model is also included.

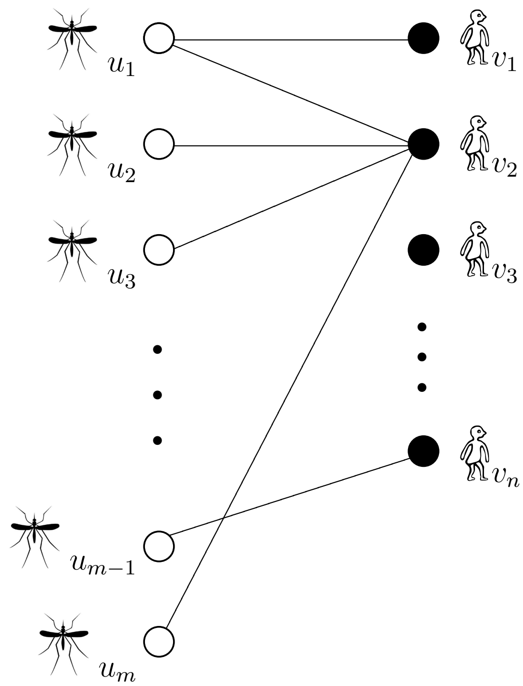
1.2 Motivating example: malaria transmission

Many biological systems are naturally modelled by random bipartite graphs. See for example [ZS02]. We define a toy example below.

Say we are modelling a system of n mosquitoes and m people. Label the mosquitoes $u_1, u_2, \ldots, u_m$ and the people $v_1, v_2, \ldots, v_n$ and draw each as a vertex of a graph. Malaria can be spread when a mosquito (u_k) bites a person (v_j) which we denote by drawing an edge between u_k and v_j.

There are no edges between any two vertices u_i and u_j because there is human to human[2] transmission of malaria (and similarly no mosquito to mosquito transmission ensures no edges between the v_j and v_k). This ensures that the graph drawn to convey this information will be bipartite. The requirement that each mosquito can only bite each person once ensures that the graph has no multiple edges.

[2]forgetting blood transfusions



Here, s_j, the degree of u_j, is the number of people bitten by the mosquito, u_j, and t_k denotes the number of bites person v_k received. We have defined our bipartite graph representation of the malarial model and now can detail the interpretation of our random graph models: $\mathcal{G}_p, \mathcal{G}_M, \mathcal{G}_{\boldsymbol{t}}$. We also illustrate the malarial interpretation of $\mathcal{G}_{\boldsymbol{t}}$ with a detailed calculation.

In the following three examples a bipartite graph is chosen at random from the bipartite graph models $\mathcal{G}_p$, $\mathcal{G}_M$ and $\mathcal{G}_{\boldsymbol{t}}$ respectively. Suppose we interpret the random graph generated as the bite pattern observed after a night with m mosquitos and n people left in the same room. We list the assumptions about the spread of malaria implicit in each model.

Malaria interpretation of the graph p-model $\mathcal{G}_p$.
Assumption: Between each mosquito and human there is an equal and independent probability, p, of a bite occurring.

Malaria interpretation of the graph *edge*-model $\mathcal{G}_M$.
Assumption: Overnight there are M bites altogether. Between each mosquito and human there is an equal likelihood of a bite occurring.

Malaria interpretation of the graph *half*-model $\mathcal{G}_{\boldsymbol{t}}$.
Assumption: Overnight the number of bites each person received is recorded in an or-

dered[3] n-tuple $\boldsymbol{t} = (t_1, t_2, \ldots, t_n)$. All combinations of particular mosquitoes causing bites to the humans that yield the recorded bites (i.e. the correct $\boldsymbol{t}$) are equally likely.

Suppose we are interested in the number of bites made by each of the mosquitoes. For example we may want to know the likelihood that one mosquito makes greatly many more bites than the other mosquitoes. In this case it would become very critical whether or not this highly active mosquito was infected with malaria. We record the number of bites each mosquito receives in a m-tuple $\boldsymbol{s} = (s_1, s_2, \ldots, s_m)$. Then the probability that all mosquitoes make the number of bites encoding in $\boldsymbol{s}$ given that the people were bitten as in $\boldsymbol{t}$ is precisely $\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(\boldsymbol{s})$.

Toy Example: Consider a system with five people in a room overnight with two mosquitos. The people, in alphabetical order, were bitten 1,1,2,2 and 1 times. It is unknown which mosquitoes caused these bites.

We assume that any scenario, under which the people were bitten the prescribed number of times, is equally likely. That is, we assume that any bipartite graph with a degree sequence of (1,1,2,2,1) for the *black* vertices is equally likely. We draw each of these bipartite graphs below and list the degree sequence of the *white* vertices for each graph.

We find the probability that the first mosquito, u_1, makes 3 bites and the second mosquito, u_2 makes 4 bites. i.e. $\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(\boldsymbol{S} = (3,4))$.

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}\big(\boldsymbol{S} = (3,4)\big) = \mathbb{P}_{\mathcal{G}_{M=7}}\big(\boldsymbol{S} = (3,4) \mid \boldsymbol{T} = (1,1,2,2,1)\big) = \tfrac{3}{8}$$

Similarly the probability that the mosquitos make 2 and 5 bites respectively is,

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}\big(\boldsymbol{S} = (2,5)\big) = \mathbb{P}_{\mathcal{G}_{M=7}}\big(\boldsymbol{S} = (2,5) \mid \boldsymbol{T} = (1,1,2,2,1)\big) = \tfrac{1}{8}$$

[3]i.e. t_k is the number of bites received by the k-th person, v_k.

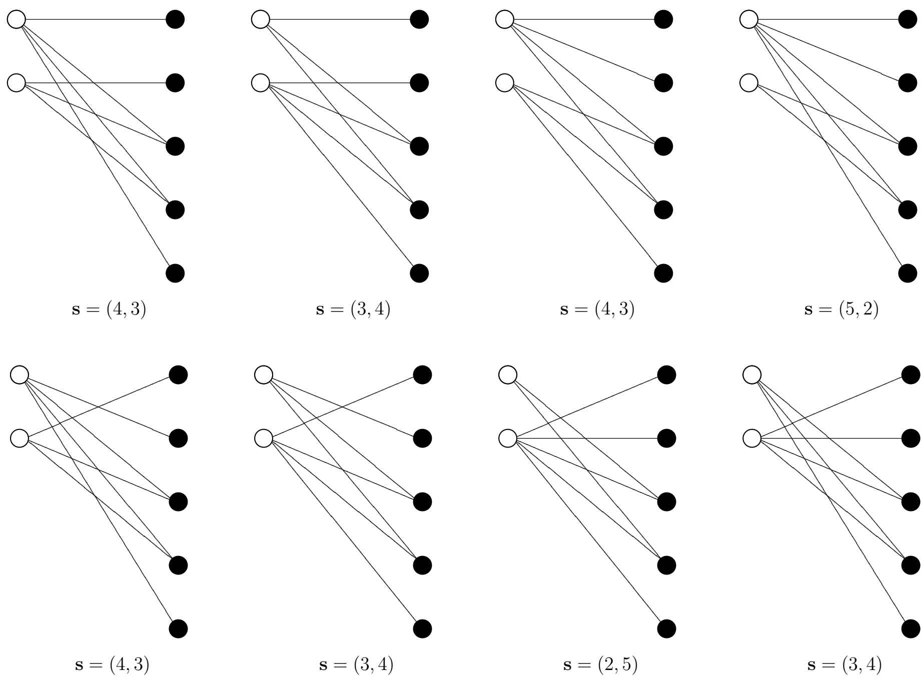


Figure 1.2: All bipartite graphs on $(2, 5)$ vertices with black degree sequence $(1, 1, 2, 2, 1)$

Chapter 2

Literature review

This thesis concerns the asymptotic behaviour of random bipartite graphs.

The general field of random graphs originated with an influential paper of Erdős and Rényi, [ER61] (see Section 2.1.1). One of the key properties introduced in [ER61] was that small changes in the initial parameters could lead to abrupt changes in the asymptotic behaviour of random graphs. This behaviour has been likened to phase transitions in physics and hence began the interest in what are termed *threshold functions* in random graphs. Threshold functions are discussed in 2.1.2.

The particular aspect of random graphs that we consider is the asymptotic behaviour of vertex degrees. Results in this area for both the general and bipartite cases are (briefly) surveyed in Sections 2.1.3 and 2.2.1 respectively.

We also compile some graph theory results counting the asymptotic number of graphs with a particular degree sequence. Enumeration results have been used to construct binomial models which approximate the general random graph models in [MW97]. Some results from this paper are described in Section 2.1.4. This thesis could could be considered a bipartite analogue of [MW97]. We proceed from a bipartite graph enumeration result to show that the asymptotic probabilities of degree sequences in random bipartite graphs can be approximated by binomial models. The bipartite enumeration result we use is from [GM09]. We give details of this result in Section 2.2.2.

The material in this chapter comes under three general headings: analogous results, previous work not used in this thesis and previous work on which this thesis builds. The only section which we will directly make use of is Section 2.2.2 in which we give results on the asymptotic enumeration of bipartite graphs. Section 2.1.4 covers a result in general graph theory which is analogous to our work. The remainder of the literature review will not be needed in the thesis but serves to give context for our results.

2.1 General random graphs

2.1.1 Early results

The field of random graphs began with two influential papers of Erdős and Rényi, [ER61] and [ER59].

A random graph. To define a random graph it is necessary to construct a domain from which it is sampled and a way in which it is sampled from that domain. For the traditional non-bipartite case, two models, $\hat{\mathcal{G}}_p(n)$ and $\hat{\mathcal{G}}_M(n)$, are the most extensively studied. (These models are analogous to $\mathcal{G}_p(m,n)$ and $\mathcal{G}_M(m,n)$ which we will later define for bipartite graphs. The hat over the $\mathcal{G}$ in our notation is not usual but used here to avoid confusion between the general case and the bipartite case.)

There is an interesting construction introduced by Janson in [Jan94] which allows us to define both models ($\hat{\mathcal{G}}_p(n)$ and $\hat{\mathcal{G}}_M(n)$) at the same time. Begin with n labelled vertices. Let $t_1, \dots, t_{\binom{n}{2}}$ be times distributed uniformly at random in $(0,1)$, then labelled in ascending order. We now add an edge at random at each time t_i. This yields a random graph in $\hat{\mathcal{G}}_p(n)$ at time $p \in (0,1)$ but also a random graph in $\hat{\mathcal{G}}_M(n)$ at the M^{th} time, t_M.

When Erdős and Rényi's paper [ER59] was published, an already famous topic in traditional (non-random) graph theory was the *chromatic number* of graphs.

Chromatic number, χ. The *chromatic number* of a graph G, denoted $\chi(G)$, is defined to be the number of colours required to colour the vertices of G in such a way that any two vertices connected by an edge in G are coloured differently.

Theorem 2.1 (Erdős and Rényi [ER61]). *Let G be a random graph in $\hat{\mathcal{G}}_M(n)$ with parameter $M = M(n)$. Fix positive constants $c < \frac{1}{2}$ and k. Then as $n \to \infty$ the following results hold almost surely,*

$$\begin{aligned} &\chi(G) &&= 2 &&\textit{if} \quad M(n) = o(n), \\ &\chi(G) &&= 3 &&\textit{if} \quad M(n) = cn, \\ \textit{and} \qquad &\chi(G) &&> \tfrac{n}{k} &&\textit{if} \quad M(n) = \tbinom{n}{2} - o(n^{2(1-1/k)}). \end{aligned}$$

This would be the first result on the chromatic number of a random graph. Much later, it was shown by [SS87] and [Łuc91] that the chromatic value of a random graph G in $\hat{\mathcal{G}}_p$ is almost surely within one of the expected chromatic value over all graphs in $\hat{\mathcal{G}}_p$. This is a seminal result in random graph theory. The paper [SS87] is also of particular interest to us as it applies a powerful probabilistic technique, *Doob's martingale process*. A similar application of this technique is central to our proof that *pathological* degree sequences are rare in the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$. We give an indication of the proof in [SS87] in Section 3.2.4.

Notice that in Theorem 2.1, varying the value of the parameter $M(n)$ changed the likely chromatic number of a graph picked at random from $\hat{\mathcal{G}}_p$. This is an example of a threshold function on random graphs, a class of functions introduced in [ER61].

2.1.2 Threshold functions

Erdős and Rényi's paper [ER61] sparked a flurry of papers deriving the threshold functions for different properties of random graphs. Examples include ....
These many isolated results were set in a general framework by Bollobás and Thomason in [BT87]. They gave a non-constructive proof showing the existence of threshold functions for a wide class of properties on graphs. We give this result below in Theorem 2.2, but prior to this we must make some definitions.

In this section we work in the probability space $\hat{\mathcal{G}}_M(n)$ where every graph on n vertices with M edges is equally likely. Consider this space for different values of the parameter $M(n)$. Let $\mathcal{G}(n)$ be the set of all labelled graphs on n vertices.

A property Q is *monotone increasing* on $\mathcal{G}(n)$ if graph $A \in \mathcal{G}(n)$ satisfying Q implies that any graph $B \in \mathcal{G}(n)$ for which A is a subgraph must also satisfy Q. A *non-trivial* property is one such that the set of graphs $G \in \mathcal{G}(n)$ for which the property holds as well as the set of graphs $G' \in \mathcal{G}(n)$ for which the property fails are both non-empty. For example, the property of having a triangle as a subgraph is a *non-trivial monotone increasing* property on the set $\mathcal{G}(n)$ for $n \geq 3$.

Notice that if a property is non-trivial and monotone increasing then Q must hold for the complete graph and fail for the empty graph. Hence we have $\mathbb{P}_{\hat{\mathcal{G}}_{M=0}}(Q) = 0$ and $\mathbb{P}_{\hat{\mathcal{G}}_{M=\binom{n}{2}}}(Q) = 1$. This is important as it guarantees the existence of the function $M^*(n)$ in Theorem 2.2. (The expression, $\mathbb{P}_{\hat{\mathcal{G}}_M}(Q)$, is the probability that a graph chosen at random from those with M edges will have property Q. The hat over the $\mathcal{G}$ is to indicate that we are working in the space of general random graphs and not random bipartite graphs.)

We are now ready to state the Bollobás and Thomason's Theorem showing general existence of threshold functions. This appeared as Theorem 4 in their 1985 paper [BT87]. Their original theorem is for more general sets, but we give the result only as it applies to random graphs in $\hat{\mathcal{G}}_M$.

Theorem 2.2 (Bollobás & Thomason 85)**.** *Let Q be a monotone increasing property on the set of all labelled graph on n vertices, $\mathcal{G}(n)$. Define the function* $M^*(n) = \max\{l :$

$\mathbb{P}_{\hat{\mathcal{G}}_{M=l}}(Q) \leq \frac{1}{2}\}$ *and let* $w(n) \to \infty$.

Then for any $M(n) \leq M^*(n)/w(n)$,

$$\mathbb{P}_{\hat{\mathcal{G}}_M}(Q) \leq 1 - 2^{-1/w},$$

and for any $M(n) \geq (M^*(n)+1)w(n)$,

$$\mathbb{P}_{\hat{\mathcal{G}}_M}(Q) \geq 1 - 2^{-w}.$$

In the theorem the function $M^*(n)$ is a threshold function for Q. Hence to show the existence of a threshold function for any property by this theorem it is sufficient to show that the property is non-trivial and monotone increasing. In many cases this is not hard[1].

2.1.3 Vertex degrees

There are two excellent surveys of this area, chapter 3 of [Bol01] and section two on [Pal88].

In the subject of vertex degrees in general random graphs it is the behaviour of two random variables that are the most studied. We define these below,

> Let $X^{(r)}$ be the random variable that returns the r^{th} largest degree in the graph and let ζ_r be the random variable that returns the number of vertices of degree r.

Number of vertices of degree r. We will first consider the random variable ζ_r. For a fixed r, the likely value of ζ_r is dependent on the expectation of ζ_r. The following result appears in [Bol01] and is from Bollobás' paper [Bol82].

By *almost surely* we mean with probability tending to one as $n \to \infty$.

Lemma 2.3 (Bollobás 82)**.** *Fix* $t \in \mathbb{R}^+$ *and* $\varepsilon > 0$. *Let* $\hat{\mathcal{G}}_p = \hat{\mathcal{G}}_p(n)$ *where* $p(1-p) \geq \varepsilon n^{-3/2}$. *Then almost surely,*

$$\begin{array}{lll} \zeta_r = 0 & \text{if} & \mathbb{E}_{\hat{\mathcal{G}}_p}(\zeta_r) \to 0 \\ \zeta_r \geq t & \text{if} & \mathbb{E}_{\hat{\mathcal{G}}_p}(\zeta_r) \to \infty. \end{array}$$

Also, if $\mathbb{E}_{\hat{\mathcal{G}}_p}(\zeta_r) = c$,

$$\forall 0 \leq k \leq n, \qquad \mathbb{P}_{\hat{\mathcal{G}}_p}\big(\zeta_r = k\big) \to \frac{1}{k!}e^{-c}c^k.$$

[1]For example consider the property of having a particular non-empty subgraph H on k vertices for any $k \leq n$. This property is non-trivial and monotone increasing on the set $\mathcal{G}(n)$. (Why? Observe H *is not* a subgraph of the empty graph and *is* a subgraph of the complete graph so the property is non-trivial. Also, if H is a subgraph of the graph A and A is a subgraph of B then B contains all the edges of A and so H must be a subgraph of B. Hence the property is also monotone increasing and so we are done.)

Note that this last results is often written as $\zeta_r \sim \mathrm{Po}(c)$ where $\mathrm{Po}(c)$ denotes the Poisson distribution with parameter c. Results similar to Lemma 2.3 hold for the distributions of the number of vertices with degree *at least* r and also for the number of vertices with degree *at most* r (see [Bol01][p.63]).

The r^{th} largest degree. Let $d_{(r)}$ be the random variable which takes the value of the r^{th} largest degree. Naturally two particular values of r are the most studied, namely $d_{(1)}$ the maximum degree and $d_{(n)}$ the minimum degree.

For suitable p these minimum and maximum degrees are unique. This was shown by Erdős and Wilson in [EW77]. The following theorem appears in [Bol01].

Theorem 2.4 (Erdős and Wilson 1977)**.** *Let $\hat{\mathcal{G}}_p = \hat{\mathcal{G}}_p(n)$ for some $p < \frac{1}{2}$ such that $\frac{pn}{\ln n} \to \infty$. Let A be the event that a graph has both a unique minimum and a unique b maximum degree. Then as $n \to \infty$,*

$$\mathbb{P}_{\hat{\mathcal{G}}_p}(A) \to 1.$$

The probability that the maximum degree is unique depends on the value of the parameter p. For example, if $p = o(\frac{1}{n} \ln n)$ then almost surely neither the maximum nor minimum degree is unique [p.68][Bol01].

There are results which show that sometimes this almost sure uniqueness can be extended beyond the maximum and minimum degrees. Fix i. Let B_i be the event that the largest i degrees are all unique, i.e. $d_{(n-i+1)} < \ldots < d_{(n)}$. Palka, in Lemma 2.1.2 of [p.17][Pal88] provides conditions under which $\mathbb{P}_{\hat{\mathcal{G}}_p}(B_i) \to 1$ as $n \to \infty$.

2.1.4 Approximation by binomial models

The paper by McKay and Wormald [MW97] constructs random models based on binomial variables that approximate the distribution of degree sequences in random graphs as the number of vertices gets arbitrarily large. This paper is of particular interest to us as their results, which concern general random graphs, are analogous to the results we prove for bipartite random graphs.

We describe the enumeration result, Theorem 2.5, which forms the starting point for their paper [MW97].

Enumeration results Havel and Hakimi independently found the conditions under which an n-tuple is the degree sequence of a graph on n vertices in [Hav55] and [Hak62] respectively. In lieu of a simple closed form solution, much work has been done finding asymptotic results for the number of graphs with given degree sequence. We are interested in the following enumeration result, also by McKay and Wormald, in [MW90]. (Let $|G(\boldsymbol{d})|$ denote the number of graphs on n vertices $v_1, \ldots, v_n$ such that the degree of each v_i is d_i. Also, write $\lambda = \binom{n}{2}^{-1} \sum_i d_i$ and $d = \frac{1}{n} \sum_i d_i$.)

Theorem 2.5 (McKay & Wormald 90)**.** *Suppose $\boldsymbol{d}$ is an n-tuple such that for each $0 \leq i \leq n$, $|d_i - d| \leq n^{1/2+\varepsilon}$ and $\lambda, 1-\lambda \geq \frac{c}{\log n}$ for some $c > \frac{4}{3}$. Let $\gamma = (n-1)^{-2} \sum_{i=1}^{n} (d_i - d)^2$. Then,*

$$|G_n(\boldsymbol{d})| = \sqrt{2} \prod_{i=1}^{n} \binom{n-1}{d_i} \exp\Big(\frac{1}{4} - \frac{\gamma^2}{4\lambda^2(1-\lambda)^2} + o(1)\Big).$$

Binomial models McKay and Wormald construct binomially based models which will later be shown to approximate the random graph models closely. Each model is based on independent binomially distributed random variables with parameters $(n-1, p)$ subject to certain constraints.

Observe that in $\hat{\mathcal{G}}_p$ if one considers a particular vertex v_i then there are $n-1$ other vertices and an independent probability p of joining to each one. Thus the degree of v_i is binomially distributed with parameters $(n-1, p)$. We note however that the degrees of two different vertices v_i and v_j are not independent; for example if the degree of v_i is $n-1$ then there must be an edge between v_i and v_j and so the degree of v_j is at least 1. Also the sum of all the degrees is twice the number of edges in the graph, so in particular the sum of the degrees must be even. This motivates the first definition.

Let $\hat{\mathcal{E}}_p$ be the space of n binomially distributed random variables with parameters $(n-1, p)$ subject to even sum.

McKay and Wormald show in Lemma 2.2 [MW97] that,

$$\mathbb{P}_{\mathcal{E}_p}(\boldsymbol{d}) = \left(\frac{1}{2} + \frac{1}{2}(q-p)^{2N}\right)^{-1} p^m q^{2N-m} \prod_{i=1}^{n} \binom{n-1}{d_i}^2.$$

Consider $\hat{\mathcal{G}}_M$, the probability space in which every graph with M edges has equal probability. The aim is to construct $\hat{\mathcal{E}}_M$ to approximate $\hat{\mathcal{G}}_M$.

Let $\hat{\mathcal{E}}_M$ be the space of n binomially distributed random variables with parameters $(n-1, p)$ subject to sum M.

Also define an integrated model over n-tuples with even sum,

$$\mathbb{P}_{\hat{\mathcal{I}}_p}(\boldsymbol{d}) = \frac{2}{V(p)\big(1-(q-p)^{2N}\big)} \prod_{i=1}^{n} \binom{n-1}{d_i} \int_0^1 K_p(p')(p')^m(1-p')^{2N-m}\,dp',$$

$$\text{where} \qquad K_p(p') = \sqrt{\frac{N}{\pi pq}}\exp\Big(-\frac{(p-p')^2N}{pq}\Big) \qquad \text{and} \qquad V(p) = \int_0^1 K_p(p')\,dp'.$$

Results These binomial models are shown to provide good approximations to the distribution of the degree sequences in the general random graph models $\hat{\mathcal{G}}_p$ and $\hat{\mathcal{G}}_M$. McKay and Wormald original theorem in [MW97] applies to all random variables on any normed space but we state it here only for real random variables. (Let $\hat{I}_n$ be the set of n-tuples with entries between 0 and $n-1$ and having even sum. The theorem holds only for what McKay and Wormald term *acceptable* values of the parameters p, M, see [MW97] for the definition.)

Theorem 2.6 (McKay and Wormald [MW97]). *For $n \geq 1$, let $X_n : \hat{I}_n \to \mathbb{R}$ be a random variable. Let $\omega(n)$ be any function such that $\omega(n) \to \infty$ and $\varepsilon(n)$ be any function such that $\varepsilon(n) \to 0$.*

Then as $n \to \infty$ subject to $p = p(n)$ is acceptable,

$$|\mathbb{E}_{\hat{\mathcal{G}}_p}(X_n) - \mathbb{E}_{\hat{\mathcal{I}}_p}(X_n)| = o(1)\mathbb{E}_{\hat{\mathcal{I}}_p}(|X_n|) + O\Big(n^{-\omega(n)} + \exp(-\varepsilon(n)(pqN)^{1/3})\Big) \max_{\boldsymbol{d}\in\hat{I}_n} |X_n(\boldsymbol{d})|.$$

Similarly, as $n \to \infty$ subject to $M = M(n)$ is acceptable,

$$|\mathbb{E}_{\hat{\mathcal{G}}_M}(X_n) - \mathbb{E}_{\mathcal{E}_M}(X_n)| = o(1)\mathbb{E}_{\mathcal{E}_M}(|X_n|) + n^{-\omega(n)} \max_{\boldsymbol{d}\in I_{n,m}} |X_n(\boldsymbol{d})|.$$

The proof of this result (in [MW97]) involves showing that,

$$\mathbb{P}_{\hat{\mathcal{G}}_p}\left(\frac{1}{4} - \frac{\gamma^2}{4\lambda^2(1-\lambda)^2} = o(1)\right) \tag{2.1}$$

is asymptotically very close to 1. The importance of this relates to the enumeration result in Theorem 2.5. If the value of (2.1) is close to 1 then this means that for 'most' degree sequences in $\hat{\mathcal{G}}_p$, Theorem 2.5 provides a simple asymptotic enumeration for the number of graphs with that degree sequence.

We will show the bipartite analogue of Theorem 2.6 in Theorems 7.4 and 7.5.

2.2 Bipartite random graphs

Random bipartite graphs relate to another class of random graphs. Let us refer to the graph on three vertices with every edge present as a triangle. A graph which does not contain this graph as a subgraph is called *triangle free*. Asymptotically almost all triangle free graphs are bipartite [EKR76]. Also, as bipartite graphs contain no odd cycles, all bipartite graphs are triangle free. Hence results concerning triangle free graphs can imply results on bipartite graphs and vice-versa.

There are many results for random bipartite graphs concerning what are called 'matchings'. A *matching* of a graph is a subset, E', of the edge set such that each vertex in the graph is incident with at least one edge in the set E'. These matching results on random bipartite graphs are surveyed in []. Janson et. al. use results on matchings in random bipartite graphs to derive similar results in general random graphs, see [JŁR00, p.85]. However, as matchings do not directly concern vertex degrees we will not include these results in this section.

The work we shall focus on is divided into two sections, the first concerns the vertex degrees in random bipartite graphs and the second is on enumeration results which count bipartite graphs by degree sequence.

We include Section 2.2.1 on vertex degrees because it is this field of work in which our thesis fits. Thus it is important to survey previous results in this area to a give context to our work.

The enumeration results in Section 2.2.2 are included for a vastly different reason. In this thesis we will approximate the probability of finding a particular degree sequence $(\boldsymbol{s}, \boldsymbol{t})$, in each of our random graph models. The asymptotic enumeration result forms the basis for these calculations.

2.2.1 Vertex Degrees

We consider the properties of the degrees of the vertices in random graphs. This is the body of work into which our results will fit so we make a brief survey of previously known results in the area. Many results steam from results in the general random graphs that can also be applied to the bipartite case. Except for the preliminary results none of these will be needed in the thesis so this section is entirely a literature survey.

There are very few results on the vertex degrees of random bipartite graphs. Many of the results that have been found stem from similar results on the vertex degrees in general random graphs. Thus the results in this section can seem somewhat disjointed as they mostly follow from their counterparts in the general case rather than having been derived from other results in bipartite random graph theory. For completeness, we include all results known to us in the area of asymptotics of vertex degrees in random bipartite graphs.

Preliminary Consider the bipartite graph p-model, $\mathcal{G}_p(m,n)$. Fix a *white* vertex u_j and let S_j be the random variable that returns the degree of u_j. By Definition 1.6 of $\mathcal{G}_p$, there is an independent probability p that u_j is joined to each of the n *black* vertices. Hence, as noted in [Pal84],

$$\forall 0 \le r \le n, \qquad \mathbb{P}_{\mathcal{G}_p}\big(S_j = r\big) = \binom{n}{r} p^r q^{n-r}. \tag{2.2}$$

We say that S_j is binomially distributed with parameters n and p. Let ξ_r be the random variable which returns the number of *white* vertices with degree r. By (2.2), each of the m *white* vertices has degree r with probability $\binom{n}{r}p^r q^{n-r}$. Hence, ξ_r is binomially distributed with parameters m and $\binom{n}{r}p^r q^{n-r}$. By this we mean,

$$\forall 0 \le k \le n, \qquad \mathbb{P}_{\mathcal{G}_p}\big(\xi_r = k\big) = \binom{m}{k}\left(\binom{n}{r}p^r q^{n-r}\right)^k \left(1 - \binom{n}{r}p^r q^{n-r}\right)^{m-k}. \tag{2.3}$$

Result (2.3) appears in [Pal84]. Palka also notes that the converse is true for η_r, the number of *black* vertices with degree r. This random variable, η_r, is binomially distributed with parameters n and $\binom{m}{r}p^r q^{m-r}$.

Number of vertices of degree r Godbole et. al. in [GLS99] study random bipartite graphs in which the numbers of *black* and *white* vertices are equal (in our notation, $m = n$). They consider an $n \times n$ chessboard in which a rook is placed on each square with independent probability $p(n)$. There is a natural bijection between this *random chessboard* and a random graph in $\mathcal{G}_p(n,n)$, where a rook in row j column k corresponds to an edge between vertices u_j and v_k. The degree of u_j (resp. v_k) then corresponds to the number of rooks in row j (resp. column k). Note any rooks in the same column or same row can be considered *mutually threatening*. Godbole et. al. define the random variable Y_r to be the number of sets of r mutually threatening rooks. We observe that this corresponds to the number of vertices with degree $h \ge r$, weighted by $\binom{h}{r}$. Let ζ_r be the random variable which returns the number of vertices (either *black* or *white*) of degree r. Then Y_r can be written,

$$Y_r = \zeta_r + (r+1)\zeta_{r+1} + \tbinom{r+2}{r}\zeta_{r+2} + \dots \tbinom{n}{r}\zeta_n. \tag{2.4}$$

We translate Theorem 2.1 of [GLS99] into our own notation.

Theorem 2.1 (Godbole et. al. [GLS99]). *Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.6. Suppose also that $m = n$. Then as $n \to \infty$,*

$$\mathbb{P}_{\mathcal{G}_p}\big(Y_r = 0\big) = \begin{cases} 1 & \text{for } p = o\Big(\frac{1}{n^{1+\frac{1}{r}}}\Big) \\ 0 & \text{for } p^{-1} = o(n^{1+\frac{1}{r}}) \end{cases} \tag{2.5}$$

There is an alternate way to generate a random chessboard; place M rooks onto the board such that all $\binom{n^2}{M}$ possible arrangements are equally likely. This corresponds to the bipartite model $\mathcal{G}_M(m,n)$ for $m = n$. Godbole et. al. prove a result in this model (i.e. $\mathcal{B}_M$) similar to Theorem 2.2.

The paper also gives the following result on the limit of the distribution of Y_r.

Theorem 2.2 (Godbole et. al. [GLS99]). *Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.7 where $p = o(n^{-1})$. Suppose also that $m = n$. Then as $n \to \infty$,*

$$\mathbb{P}_{\mathcal{G}_p}\big(Y_r = k\big) = \frac{e^{-c}c^k}{k!} \qquad \text{where } c = 2n\tbinom{n}{r}p^r \tag{2.6}$$

Number of vertices with fixed degree r. So the distribution of ζ_r converges to the Poisson distribution when the expectation of ζ_r is finite and converges to the normal distribution when the expectation of ζ converges to infinity, as $o(m^\varepsilon)$. This has some interesting corollaries. The parameters in the probability distributions of ζ_r can be calculated explicitly for particular values of p. This is done by Palka [Pal84] in the following corollaries. (We write $X \sim N(0,1)$ to denote that the distribution of the random variable X converges to the normal distribution with expectation 0 and variance 1, see [Pal84] for precise definition.)

Corollary 2.3. *Fix $0 < \beta < \infty$. Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.6. Suppose also that $m = \alpha n$.*

1. *Let $r \geq 2$. Suppose $np = \beta m^{-1/r}$. Then as $m \to \infty, \zeta_r \sim P(\lambda)$.*

$$\text{i.e.} \qquad \mathbb{P}_{\mathcal{G}_p}(\zeta_r = c) = \frac{\lambda^c e^{-\lambda}}{c!} \qquad \text{where } \lambda = \frac{1}{r!}\beta^r(1-\alpha^{1-r})$$

2. *Let $r \geq 0$. Suppose $np = \ln m - \beta \ln\ln m + o(\ln\ln m)$. Then as $m \to \infty$,*

$$\frac{(\zeta_r - a)}{\sqrt{a}} \sim N(0,1) \qquad \text{where } a = \begin{cases} \frac{2}{r!}(\ln m)^{r+\beta} & \text{if } \alpha = 1 \\ \frac{1}{r!}(\ln m)^{r+\beta} & \text{if } 0 < \alpha < 1 \end{cases}$$

3. *Let $r \geq 1$ and $0 < y < 1$. Suppose $np = \ln m - (1-\gamma) r \ln\ln m + o(\ln\ln m)$. Then as $m \to \infty$,*

$$\frac{(\zeta_r - a)}{\sqrt{a}} \sim N(0,1) \qquad \textit{where } a = \begin{cases} \frac{2}{r!}(\ln m)^{r\gamma} & \textit{if } \alpha = 1 \\ \frac{1}{r!}(\ln m)^{r\gamma} & \textit{if } 0 < \alpha < 1 \end{cases}$$

4. *Let $r \geq 0$ and $-\infty < \gamma < \infty$. Suppose $np = \ln m - r\ln\ln m + \gamma + o(1)$. Then as $m \to \infty$,*

$$\zeta_r \sim P(\lambda) \qquad \textit{where } \lambda = \begin{cases} \frac{2e^{-\gamma}}{r!} & \textit{if } \alpha = 1 \\ \frac{e^{-\gamma}}{r!} & \textit{if } 0 < \alpha < 1 \end{cases}$$

5. *Let $r \geq 0$. Suppose $np = \ln m - r \ln\ln m + f(m)$, where $f(m) = o(\ln\ln m)$ tends to ∞. Then as $m \to \infty$,*

$$\mathbb{P}_{\mathcal{G}_p}\big(\zeta_r = 0\big) \to 1$$

Let ξ_r (resp. η_r) be the number of *white* (resp. *black*) vertices of degree r. The final result in [Pal84] concerns ξ_r and η_r for the case $m = n$, i.e. when the number of *white* vertices is equal to the number of *black* vertices. This result, shown in Theorem 2.4, gives the asymptotic distribution of ξ_r and η_r individually (rather than for the random variable $\zeta_r = \xi_r + \eta_r$).

Theorem 2.4. *Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.6. Suppose also that $m = n$. Then,*

$$\lim_{m\to\infty} \mathbb{P}_{\mathcal{G}_p}\big(\xi_r = i, \eta_r = j\big) = \frac{\mu^{i+j}}{i!j!} e^{-2\mu} \qquad \textit{where } \mu = e^{-\gamma}/r!$$

The limiting distribution of ξ_r and η_r is the bi-variate Poisson distribution with parameter μ [Pal84].

In a bipartite graph on (m,n) vertices the relative sizes of m and n are important. Palka, in [Pal87] shows that in $\mathcal{G}_p(m,n)$ that when the number of *white* vertices is sufficiently large in comparison to the number of *black* vertices then the vertex of maximum degree is almost surely a *black* vertex. Indeed for suitable p and suitable ratio of m greater then n, Palka shows for fixed i that almost surely the smallest i degrees will all be on *white* vertices. Similar results are shown by Rucinski for k-partite[2]graphs in [Ruc81].

In a bipartite graph the number of edges incident with each colour class of vertices is the same so the colour with fewer vertices will have a higher average degree, and so it makes sense that it is also more likely to host the vertex of maximum degree.

[2] A k-partite graph is one in which the vertices are partitioned into k sets, $V_1, \ldots, V_k$ and the edge set E is a subset of $\{(v_a, v_b) \;:\; v_a \in V_i,\;\; v_b \in V_b \;\text{ for } i \neq j\}$.

2.2.2 Enumeration results

We are interested in results that give counts for the number of bipartite graphs with a given degree sequence. Much of the early enumeration work was done in the context of binary matrices with given row and column sums. Ryser [Rys63] found the necessary and sufficient conditions on the row and column sums that guarantee the existence of a binary matrix with such sums. A closed form solution for the precise number of bipartite graphs with given degree sequence is unknown but asymptotic results are given by [Bar10] and others. We shall use the result in [GM09].

McKay and Greenhill derived a formula for finding the number of bipartite graphs for some given degree sequences Theorem 2.1 of [GM09, p.4]. This result forms the foundation of the thesis. Before stating the theorem we define (ε, a)-*regular* degree sequences.

Definition 2.5 (*acceptable* for a, m and n)**.**
Let

$$f_{m,n}(x) := \frac{(1-2x)^2}{4x(1-x)}\left(1 + \frac{5m}{6n} + \frac{5n}{6m}\right).$$

If $f_{m,n}(x) < a \ln n$ *then* x *is* acceptable *for* a, m *and* n*.*

Conditions 2.6. *These conditions apply to an* m*-tuple* $\boldsymbol{s} = \boldsymbol{s}(m) = (s_1, ..., s_m)$ *and an* n*-tuple* $\boldsymbol{t} = \boldsymbol{t}(n) = (t_1, ..., t_n)$*. The conditions depend on the parameters* a *and* ε*.*

For $m, n \to \infty$*,*

- $s_j - s$ *and* $t_k - t$ *are uniformly* $O(n^{1/2+\varepsilon})$ *for* $1 \leq j \leq m$ *and* $1 \leq k \leq n$*.*
- $\sum s_j = \sum t_k$ *and* $\lambda = \frac{1}{mn}\sum_j s_j = \frac{1}{mn}\sum_k t_k$ *is* acceptable *for* a, m *and* n*.*

Definition 2.7 ((ε, a)-*regular*)**.** If m, n and $(\boldsymbol{s}, \boldsymbol{t})$ satisfy Conditions 2.6 then we say that $(\boldsymbol{s}, \boldsymbol{t})$ is (ε, a)-*regular.*

We now state the enumeration result by Greenhill and McKay (stated as Theorem 2.1 in [GM09, p.4]). To maintain consistency with the notation of [GM09] we denote the number of graphs with degree sequence $(\boldsymbol{s}, \boldsymbol{t})$ by $|B(\boldsymbol{s}, \boldsymbol{t})|$.

Theorem 2.8 (Greenhill and McKay)**.** *Let* $a, b \in \mathbb{R}^+$ *such that* $a + b < \frac{1}{2}$*. Then there exists* $\varepsilon = \varepsilon_0(a, b) > 0$ *such that the following holds.*

If $(\boldsymbol{s}, \boldsymbol{t})$ *is* (ε, a)-regular *then as* $m, n \to \infty$,

$$|B(\boldsymbol{s},\boldsymbol{t})| = \binom{mn}{\lambda mn}^{-1} \prod_{j=1}^{m} \binom{n}{s_j} \prod_{k=1}^{n} \binom{m}{t_k} \times \exp\Bigg(-\tfrac{1}{2}\Big(1 - \frac{\sum_j (s_j - s)^2}{\lambda(1-\lambda)mn}\Big)\Big(1 - \frac{\sum_k (t_k - t)^2}{\lambda(1-\lambda)mn}\Big) + O(n^{-b})\Bigg). \tag{2.7}$$

Greenhill and McKay also note that the error term $O(n^{-b})$ is uniform. This will be important in Section 7.3.

Note that if a degree sequence $(\boldsymbol{s}, \boldsymbol{t})$, satisfies $\Big(1 - \frac{\sum_j (s_j - s)^2}{\lambda(1-\lambda)mn}\Big)\Big(1 - \frac{\sum_k (t_k - t)^2}{\lambda(1-\lambda)mn}\Big) = O(n^{-b})$, then the enumeration in (2.7) simplifies. This motivates the following definition.

Definition 2.9 ($(a, b, m, n, \varepsilon)$*-pathological*)**.** *A degree sequence is called non-*$(a, b, m, n, \varepsilon)$-pathological *if it is both* (ε, a)-regular *and* $\Big(1 - \frac{\sum_j (s_j - s)^2}{\lambda(1-\lambda)mn}\Big)\Big(1 - \frac{\sum_k (t_k - t)^2}{\lambda(1-\lambda)mn}\Big) = O(n^{-b})$.

Corollary 2.10. *Let* $a, b > 0$ *be constants such that* $a + b < \frac{1}{2}$. *Then there exists* $\varepsilon = \varepsilon_0(a, b) > 0$ *such that the following holds.*

If $(\boldsymbol{s}, \boldsymbol{t})$ *is not* $(a, b, m, n, \varepsilon)$-pathological *then as* $m, n \to \infty$,

$$|B(\boldsymbol{s},\boldsymbol{t})| = \binom{mn}{\lambda mn}^{-1} \prod_{j=1}^{m} \binom{n}{s_j} \prod_{k=1}^{n} \binom{m}{t_k} \big(1 + O(n^{-b})\big). \tag{2.8}$$

Hence, for any non-*pathological* degree sequence we now have the simple enumeration formula in Corollary 2.10. We prove in this thesis that in each of our random graph models (under suitable parameters) almost all graphs have non-*pathological* degree sequences. In particular we show this for models $\mathcal{G}_p$, $\mathcal{G}_M$ and $\mathcal{G}_{\boldsymbol{t}}$ in Theorems 4.9, 5.2 and 6.20 respectively.

Chapter 3

Probabilistic bounds and techniques

In this thesis at many points it will be necessary to show that a random variable is very likely to take a value close to its mean. To enable us to achieve this we make use of some concentration inequalities from probability theory. They are so-called because under suitable conditions they can show that a random variable is likely to be *concentrated* about its mean.

We will often show that the sum of a large number of random variables is very likely to fall close to the mean of the sum of those random variables. We will deal with two cases. Most random variables we deal with will be independent. This allows us to use the concentration results by Hoeffding and McDiarmid (Theorems 3.1 and 3.2 respectively).

In the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$, however, we will need to be able to manipulate random variables which are not independent. For this situation we construct a martingale (weaker then independence) which allows us to use a generalised Azuma theorem (Theorem 3.8) to gain the necessary concentration results.

Our method to construct this martingale follows the famous technique known as Doob's martingale process. The theory for this technique forms Section 3.2.4 where we also run through the iconic proof by Shamir and Spencer in [SS87]. This proof shows the chromatic number in a general random graph is highly concentrated. It is a good, clean example of the Doob process invaluable to understand our application of it to the graph *half*-model which has some added complications.

3.1 Concentration inequalities for independent random variables

We use some general theorems from probability theory, which give ways to compute the likelihood that variables lie very close to their expectations. These Theorems were listed in the survey by Chung and Lu [pp.84-86][CL06]. The first of these concentration equalities is due to Chernoff and appears in his paper [Che81]. The second is done by McDiarmid in his paper [McD98].

Theorem 3.1 (Chernoff 1981). *Let $X_1, \ldots, X_n$ be independent random variables such that $\mathbb{P}(X_i = 1) = p_i$ and $\mathbb{P}(X_i = 0) = 1 - p_i$. Define $X = \sum_{i=1}^n X_i$. This implies the following bounds,*

$$\mathbb{P}(X \leq \mathbb{E}(X) - k) \leq e^{-k^2/2\mathbb{E}(X)}$$

$$\mathbb{P}(X \geq \mathbb{E}(X) + k) \leq e^{\frac{-k^2}{2(\mathbb{E}(X)+k/3)}}.$$

And hence for positive k,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq k) \leq e^{-k^2/2\mathbb{E}(X)} + e^{\frac{-k^2}{2(\mathbb{E}(X)+k/3)}} < 2e^{\frac{-k^2}{2(\mathbb{E}(X)+k/3)}}.$$

In most applications, $p_i = p$ for all i, in which case $\mathbb{E}(X) = np$.
The following Theorem is used to prove Lemma 4.4. It is quite powerful in this context because in Lemma 4.4 we have strong bounds on the difference between each random variable and it's expectation.

Theorem 3.2 (McDiarmid 1998). *Let $X_1, \ldots, X_n$ be independent random variables satisfying $|X_i - \mathbb{E}(X_i)| \leq c_i$ for $1 \leq i \leq n$. Define $X = \sum_{i=1}^n X_i$. Then,*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq k) \leq e^{\frac{-k^2}{2\sum_{i=1}^n c_i^2}}.$$

3.2 Martingales

3.2.1 Preliminaries

Definition 3.1 (σ-algebra). *A set $\mathcal{F}$ is a σ-algebra if it satisfies the following conditions,*

- $\mathcal{F} \neq \varnothing$
- $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

- $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$

Suppose we have a set $P = \{Y_i\}_{i \in I}$ which partitions Ω, then there is a natural σ-algebra on Ω corresponding the partition P. The following construction is from [CMZ09].

Definition 3.2 (σ-field induced by partition P_i)**.** *Let $\mathcal{F}_i$ be the collection of all sets which may be defined as unions of blocks in the partition P_i. Then we say $\mathcal{F}_i$ the σ-field induced by the partition P_i.*

Cooper et. al. [CMZ09] also note that for a set of partitions $\{P_i\}_{0 \leq i \leq n}$ where each P_{i+1} is a refinement P_i, the set of σ-algebras $\{\mathcal{F}_i\}_{0 \leq i \leq n}$ induced by the partitions $\{P_i\}_{0 \leq i \leq n}$ form a filter.

A martingale is defined with respect to the filter as follows [CMZ09].

Definition 3.3 (martingale with respect to filter)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filter $\{\mathcal{F}_i\}_{0 \leq i \leq n}$. Suppose that $X_0, \ldots, X_n$ are random variables such that for each $0 \leq i \leq n$, X_i is $\mathcal{F}_i$ measurable. The sequence $X_0, \ldots, X_n$ is a martingale provided that*

$$\mathbb{E}\big(X_{i+1} \mid \mathcal{F}_i\big) = X_i$$

for each $0 \leq i < n$.

Conditional expectation The expectation of a random variable X conditional on a σ-algebra can be quite generally. However, we define only the special case which we will require, when the σ-algebra is induced by a partition and Ω is a finite set. Note that in the probability space $(\Omega, \Sigma, \mathbb{P})$ where $|\Omega| < \infty$ that any random variable $Y : \Sigma \to \mathbb{R}$ can be defined by giving its value at each $\omega \in \Omega$.

Definition 3.4 (Conditional expectation, $\mathbb{E}\big(X \mid \mathcal{F}\big)$)**.** *This is defined only for the special case that the σ-algebra $\mathcal{F}$ is induced by some partition $\{Y_i\}_{i \in I}$ of Ω. We also assume that Ω is finite.*

Let $\omega \in \Omega$. Now, because $\{Y_i\}_{i \in I}$ is a partition, $\exists!\ j \in I$ such that $\omega \in Y_j$. Then,

$$\mathbb{E}\big(X \mid \mathcal{F}_i\big)(\omega) := \frac{\sum_{\omega' \in Y_j} X(\omega')\mathbb{P}(\omega')}{\mathbb{P}(Y_j)}.$$

We observe the following two special cases. Let $\mathcal{F}_0 = \{\varnothing, \Omega\}$. Then $\mathcal{F}_0$ is induced by the trivial partition $\{\Omega\}$, and,

$$\mathbb{E}\big(X \mid \mathcal{F}_0\big)(\omega) = \frac{\sum_{\omega' \in \Omega} X(\omega')\mathbb{P}(\omega')}{\mathbb{P}(\Omega)} = \mathbb{E}(X). \tag{3.1}$$

Now suppose, conversely, that we have the finest possible partition, $P = \{\{\omega\} : \omega \in \Omega\}$. Let $\mathcal{F}^*$ be the σ-algebra induced by P, then,

$$\mathbb{E}\big(X \mid \mathcal{F}^*\big)(\omega) = \frac{\sum_{\omega' \in \{\omega\}} X(\omega')\mathbb{P}(\omega')}{\mathbb{P}(\omega)} = X(\omega). \tag{3.2}$$

The following lemma, which we use to construct martingales, is a special case of the tower of expectation property. For statement and proof of the more general case refer to [Wil91, p.88].

Lemma 3.5. *Let $\{Y_i\}_{i\in I}$ and $\{Z_j\}_{j\in J}$ both be partitions of Ω with finite index sets I and J. Suppose further that for each $i \in I$, $\exists! J_i \subset J$ such that $\{Z_j\}_{j\in J_i}$ is a partition of Y_i. Let $\mathcal{F}_k$ and $\mathcal{F}_{k+1}$ be the σ-algebras induced by partitions $\{Y_i\}_{i\in I}$ and $\{Z_j\}_{j\in J_i}$ respectively, according to Definition 3.2. Then,*

$$\mathbb{E}\Big(\mathbb{E}\big(X \mid \mathcal{F}_{k+1} \mid \mathcal{F}_k\Big) = \mathbb{E}\big(X \mid \mathcal{F}_k\big).$$

3.2.2 Azuma-Hoeffding inequality

This theorem is often referred to as Azuma's inequality as it appeared in Azuma's 1967 paper [Azu67], but it also appeared earlier in Hoeffding's 1963 paper [Hoe63].

Theorem 3.6 (Azuma-Hoeffding)**.** *Suppose $X_0, \dots, X_n$ form a martingale such that $|X_k - X_{k-1}| \le c_k$ for each $1 \le k \le n$. Then,*

$$\mathbb{P}\big(|X_n - X_0| \ge r\big) \le e^{\frac{-r^2}{2\sum_{k=1}^n c_k^2}}.$$

There is also a strengthened version of Theorem 3.6 proven by McDiarmid in [McD89]. He showed that the expression in the exponent could improved by a factor of 4. (That is, the right hand side becomes $\exp\left(-2r^2/\sum_{k=1}^t c_k^2\right)$ for the same initial conditions.)

3.2.3 Generalised Azuma-Hoeffding inequality

We also define a more general version of the Azuma-Hoeffding inequality, it applies even when the bound $|X_k - X_{k-1}| \le c_k$ does not hold on the entire space. A similar result appears in [Vu02, p.9] as Lemma 3.1. It also appears as Proposition 3 in [HHV09] who give a proof that it follows from Theorem 3.6.

Definition 3.7 (near-c-Lipschitz with exceptional probability η)**.** A martingale is *near-c-Lipschitz with exceptional probability* η if,

$$\mathbb{P}(|X_i - X_{i-1}| \geq c) \leq \eta.$$

Theorem 3.8 (Azuma-Hoeffding)**.** *Suppose $X_0, \ldots, X_n$ forms a martingale which is near-c-Lipschitz with exceptional probability η. Then,*

$$\mathbb{P}\big(|X_n - X_0| \geq r\big) \leq e^{\frac{-r^2}{2nc^2}} + \eta.$$

We will use Theorem 3.8 in Lemma 6.18 as part of our proof that *pathological* degree sequences are rare in the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$.

3.2.4 Doob's martingale process

The method we will describe is often referred to as *Doob's martingale process* and is often used to show that the value of a random variable is concentrated about its mean.

Suppose we want to bound the difference between the expected value of a function f over the whole domain (C, say) and the value of the function on a specific point in the domain ($c \in C$). We set $X_0 = E[f(c)]$ and for a fixed n, $X_n = E[f(x)|x \in C]$. The method then proceeds to define a series of random variables X_i for $0 \leq i \leq n$ such that $X_i = E[f(x)|x \in C_i]$, where $c = C_0 \subset \ldots \subset C_i \subset C_{i+1} \subset \ldots \subset C_n = C$. We are in effect, *zooming in* on the value of f at c as we take the expected value of f over smaller and smaller subsets of C. The X_i's then form a martingale and concentration results such as Azuma's inequality can be used to bound $|X_0 - X_n|$.

Steps in Doob's martingale process on the probability space $(\Omega, \Sigma, \mathbb{P})$,

- create a filter $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_n$ (often via partitions of Ω.)
- define $X_i = \mathbb{E}(X \mid \mathcal{F}_i)$
- prove the X_i's form a martingale with respect to the filter (via tower property).
- bound the magnitude of successive differences $|X_i - X_{i-1}|$ for each $1 \leq i \leq n$.
- apply Azuma-Hoeffding Theorem 3.6

Shamir and Spencer used Doob's martingale process to show that the chromatic number of a random graph is very concentrated in [SS87]. This is a pertinent example for two reasons. Firstly, it was one of the first major results which used probabilistic techniques in combinatorics and piqued interest in the area. Secondly it is an example of a more general (widely used) technique called Doob's martingale process. Later in of this thesis we will use this technique to show Lemma 6.17. This will form part of our proof that *pathological* degree sequences are rare in the bipartite graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$.

We work in the probability space $\hat{\mathcal{G}}_p$ and begin by defining a filter $\{\mathcal{F}_i\}_{0\leq i\leq n-1}$ on this space. This filter is defined via the partition induced by the following equivalence classes.

i-equivalence of graphs Two graphs are i-equivalent, denoted $H \equiv_i G$ if the following condition holds.

$$\text{For all } 1 \leq x \leq i \text{ and all } 1 \leq y \leq n \text{ the edge } \{x,y\} \in H \Leftrightarrow \{x,y\} \in G.$$

That is, we require all edges which emanate from one of the first i vertices to be the same in both graphs. We illustrate this definition with an example in 3.1.

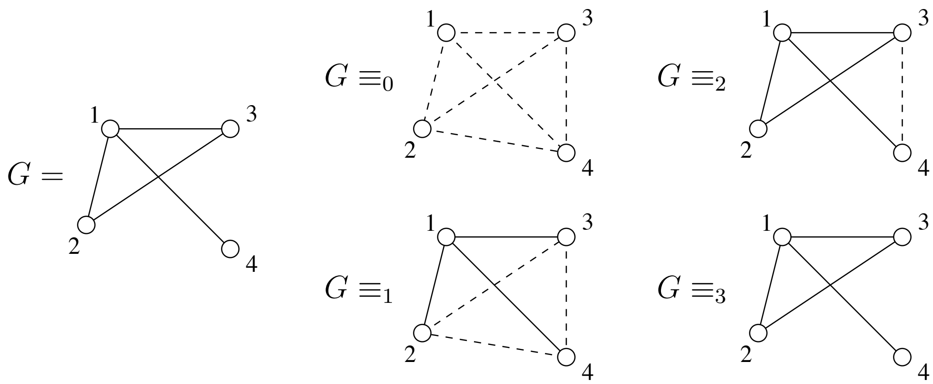


Figure 3.1: The sets of graphs i equivalent to G for $0 \leq i \leq 3$. The dashed lines indicate that edges that can be either absent or present in the graphs in that equivalence class.

For each $0 \leq i \leq n-1$ the equivalence relation defines a partition P_i. Observe that the partition P_{i+1} is a refinement of P_i. For $0 \leq i \leq n-1$ let $\mathcal{F}_i$ be the σ-algebra induced by P_i according to Definition 3.2.

Vertex uncovering martingale Given a random graph $G \in \hat{\mathcal{G}}_p(n)$, we define a sequence of random variables $X_0, \ldots, X_{n-1}$ such that $X_0(G) = \mathbb{E}_{\hat{\mathcal{G}}_p}[\chi]$ and $X_{n-1}(G) = \chi(G)$. We define,

$$X_i(G) := \mathbb{E}_{\hat{\mathcal{G}}_p}\big(\chi|\mathcal{F}_i\big)(G) = \mathbb{E}_{\hat{\mathcal{G}}_p}\big(\chi(H)|H \equiv_i G\big).$$

The first equality is our definition of X_i and the second follows by Definition 1.3 of conditional probability.

When $i = 0$, X_i the expectation of the chromatic number over all random graphs in $\hat{\mathcal{G}}_p(n)$. But when $i = n-1$, the $(n-1)$-equivalence of two graphs requires them to be the same, i.e. G is the only graph $(n-1)$-equivalent to G. Hence $X_{n-1} = \chi(G)$.

Claim The X_i's form a martingale.

Proof. This is equivalent to proving the equality $\mathbb{E}_{\hat{\mathcal{G}}_p}\Big(\mathbb{E}\big(X_{i+1} \mid \mathcal{F}_{i+1}\big) \mid \mathcal{F}_i\Big) = \mathbb{E}_{\hat{\mathcal{G}}_p}\big(X_{i+1} \mid \mathcal{F}_i\big)$. But this is precisely the *Tower Property* of Expectation (see Lemma 3.5) and thus the random variables X_i, form a martingale. □

Hence we can apply Azuma's lemma but first we need to bound $|X_{i+1} - X_i|$.

Claim $|X_{i+1} - X_i| \leq 1$.

Proof. We compare X_{i+1} and X_i. X_i is the expected number of colours needed knowing the edges of the first i vertices. The random variable X_{i+1} is the expected number of colours needed knowing both edges of the first i vertices and also the edges which join to the $(i+1)^{\text{th}}$ vertex. This only affects which colours are possible at the $(i+1)^{\text{th}}$ vertex. Hence the expected value of colours can change by at most one. So we have $|X_{i+1} - X_i| \leq 1$ as claimed. □

Concentration of Chromatic Number Result By the Azuma-Hoeffding Theorem 3.6, and because $|X_{i+1} - X_i| \leq 1$ for each $0 \leq i \leq n-2$, we have,

$$\mathbb{P}_{\hat{\mathcal{G}}_p}\big(|X_{n-1} - X_0| \geq \lambda\big) \leq e^{\frac{-\lambda^2}{2(n-1)}}.$$

Hence there is high probability that the chromatic number of G is close to the average chromatic number over all random graphs in $\hat{\mathcal{G}}_p(n)$. For example, fix $\varepsilon > 0$, then,

$$\mathbb{P}_{\hat{\mathcal{G}}_p}\big(|X_{n-1} - X_0| \geq n^{1/2+\varepsilon}\big) \leq e^{\frac{-n^{1-2\varepsilon}}{(n-1)}} \leq e^{-n^{2\varepsilon}}.$$

Thus the chromatic number of a random graph G in $\hat{\mathcal{G}}_p$ is almost surely within $n^{1/2+\varepsilon}$ of the expected chromatic number as $n \to \infty$.

Shamir and Spencer show more than this. In [SS87] they prove that for $s = 3$, a random graph G in $\hat{\mathcal{G}}_p$ (and for suitable p) there exists a function $u(n)$ such that,

$$lim_{n\to\infty}\mathbb{P}_{\hat{\mathcal{G}}_p}\big(u \leq \chi(G) \leq u + s\big) = 1. \tag{3.3}$$

This was further strengthened by Łuczak in [Łuc91] who proved (3.3) for the case $s = 1$.

3.3 Probability generating functions

Probability generating functions will be a useful tool at a couple of points within this thesis. Proofs of the statements can be found in [Wil06]. Suppose we have random variables X, Y defined on some probability space $(\Omega, \Sigma, \mathbb{P})$, for $\Omega \subseteq N$. Then we say that the power series $P(x)$ is a *probability generating function* for the random variable X if it satisfies (3.4).

$$P(x) = p_0 + p_1 x + p_2 x^2 \dots \text{ where } p_i = \mathbb{P}(X = i) \tag{3.4}$$

The expectation of X, $\mathbb{E}(X)$ may now be calculated in terms of $P(x)$.

$$\begin{aligned}\mathbb{E}[X] &= 0.p_0 + 1.p_1 + 2.p_2 + 3.p_3 + \dots \\ &= (1.p_1 + 2.p_2 x + 3.p_3 x^2 + \dots)_{x=1} \\ &= \big(\frac{d}{dx}P(x)\big)_{x=1}.\end{aligned}$$

This is sometimes useful when calculating the expectation of discrete random variables. Similarly,

$$\begin{aligned}\mathbb{E}[X^2] &= 0.p_0 + 1.p_1 + 2^2.p_2 + 3^2.p_3 + \dots \\ &= \frac{d}{dx}x(1.p_1 + 2.p_2 x + 3.p_3 x^2 + \dots)_{x=1} \\ &= \big(\frac{d}{dx}x\frac{d}{dx}P(x)\big)_{x=1}.\end{aligned}$$

Let $Q(x)$ be then probability generating function for Y. Then if X and Y are independent random variables,

$$\mathbb{E}[XY] = \Big(\big(x\frac{d}{dx}\big)P(x)\Big)\Big(\big(x\frac{d}{dx}\big)Q(x)\Big)_{x=1}.$$

Part II

New results: Degree sequences in random bipartite graphs

The main goal of this thesis is to show that in each of the random graph models the probability of a non-*pathological* degree sequence $(\boldsymbol{s}, \boldsymbol{t})$ can be asymptotically approximated by the probability of an $(m+n)$-tuple in a specially constructed binomial model.

In this part we will find an asymptotic probability for non-*pathological* degree sequences in each of the random graph models. The starting point for this is Greenhill and McKay's Theorem, our Theorem 2.8, which gives a count for the number of graphs with degree sequence $(\boldsymbol{s}, \boldsymbol{t})$. From this counting result we will derive the probability of a given degree sequence in each of our random graph models.

Theorem 2.8 does not hold for all degree sequences, only (ε, a)-*regular* degree sequences. Hence our first result, for each random graph model, is to show that these (ε, a)-*regular* degree sequences account for the bulk of the probability space.

Our second result concerns a further restriction on (ε, a)-*regular* degree sequence. Observe that if an (ε, a)-*regular* degree sequence satisfies

$$\left(1 - \frac{\sum_j (s_j - s)^2}{\lambda(1-\lambda)mn}\right)\left(1 - \frac{\sum_k (t_k - t)^2}{\lambda(1-\lambda)mn}\right) = O(n^{-b}) \tag{3.5}$$

then the function on the degrees in the exponential term of (2.7) disappears into the error term, $O(n^{-b})$. This is what we desire. For these well-behaved degree sequences $(\boldsymbol{s}, \boldsymbol{t})$, which we have called non-*pathological*, we have the simplified enumeration result, Corollary 2.10. So, our second result for each of our random graph models is that the probability of a degree sequence being *pathological* is asymptotically very low.

This part is organised as follows.

We begin by analysing the graph p-model, $\mathcal{G}_p$, in Chapter 4. In Section 4.1, we show that degree sequences are likely to be (ε, a)-*regular* in this probability space. Then in Section 4.2 we show that most degree sequences are non-*pathological*.

Both these steps are then repeated for the graph *edge*-model, $\mathcal{G}_M$ in Chapter 5. This is done by considering the graph *edge*-model as a conditional case of the graph p-model.

Lastly we show the same two results hold in the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$. In this model we shall see that the degrees of the *white* vertices are not independent as they were for the graph p-model, $\mathcal{G}_p$. Hence this case requires some more effort and we work with martingale concentration inequalities rather than the stronger concentration inequalities for independent random variables.

Chapter 4

Graph p-model, $\mathcal{G}_p$.

We work with our model of random graphs $\mathcal{G}_p$ where each edge of the complete bipartite graph $K_{m,n}$ is chosen independently with probability $p = p(m,n)$. See Definition 1.6.
In this model we will refer often to the following conditions.

Conditions 4.1. *These conditions apply to an m-tuple $\boldsymbol{s} = \boldsymbol{s}(m,n) = (s_1, ..., s_m)$ and an n-tuple $\boldsymbol{t} = \boldsymbol{t}(m,n) = (t_1, ..., t_n)$. The conditions depend on the parameters a and ε.*

For $m, n \to \infty$,

- *p is* acceptable *for a, m and n.*
- *$m = o(n^{1+\varepsilon})$ and $n = o(m^{1+\varepsilon})$.*

4.1 (ε, a)-regular degree sequences in $\mathcal{G}_p$

4.1.1 Variation in the degree of each vertex.

In this section we show that a graph, G, chosen at random in $\mathcal{G}_p$ has degree sequence $(\boldsymbol{s}, \boldsymbol{t})$, likely to satisfy the following for appropriate $\varepsilon > 0$.

$$\text{For each } 1 \le j \le m \text{ and } 1 \le k \le n,\ s_j - s \text{ and } t_k - t \text{ are } O(n^{1/2+\varepsilon}).$$

Lemma 4.1. *Fix $0 < a < \frac{1}{2}$ and $\varepsilon > 0$. Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.6.*

1. *Let S be the random variable that returns the average degree of the* white *vertices $u_1, \ldots, u_m$. Then as $m, n \to \infty$ subject to Conditions 4.1,*
$$\mathbb{P}_{\mathcal{G}_p}\big(|S - np| \ge n^{3\varepsilon}\big) \le e^{-n^{2\varepsilon}}$$

2. *For* $1 \leq j \leq m$*, let* S_j *be the random variable that returns the* degree *of the* white *vertex* u_j*. Then as* $m, n \to \infty$ *subject to Conditions 4.1,*

$$\mathbb{P}_{\mathcal{G}_p}\big(|S_j - np| \geq n^{1/2+\varepsilon/3}\big) \leq e^{-n^{3\varepsilon/2}}$$

Proof. (of 1.) We firstly observe that

$$|mS - mnp| \leq n^{1+2\varepsilon} \leq n^{3\varepsilon} m \tag{4.1}$$

implies that

$$|S - np| \leq n^{3\varepsilon}.$$

We prove (4.1) using Chernoff's inequality (Theorem 3.1).

Construct the random variable $X_{j,k}$ by setting $X_{j,k} = 1$ if there is an edge between u_j and v_k in our random graph G and otherwise setting $X_{j,k} = 0$.

Then $mS = \sum_{j,k} X_{j,k} = X$, and $\mathbb{E}_{\mathcal{G}_p}\big(mS\big) = \mathbb{E}_{\mathcal{G}_p}\big(X\big) = mnp$. By Lemma 4.5, $p > \frac{1}{\ln n}$ and so $mnp < mn^{1+\varepsilon}$. Set $k = n^{1+2\varepsilon}$, then by Chernoff's inequality (Theorem 3.1),

$$\mathbb{P}_{\mathcal{G}_p}\big(|mS - mnp| \geq n^{1+2\varepsilon}\big) \leq 2e^{\frac{-n^{2+4\varepsilon}}{2(n^{2+\varepsilon}+n^{1+2\varepsilon}/3)}} \leq e^{-n^{2\varepsilon}}$$

hence,

$$\mathbb{P}_{\mathcal{G}_p}\big(|S - np| \geq n^{3\varepsilon}\big) \leq e^{-n^{2\varepsilon}}$$

□

Proof. (of 2.) We again use Chernoff's inequality (Theorem 3.1), this time to bound the probability that S_j is concentrated about its expected value, np.

We use the random variables $X_{j,k}$ defined above. The number of edges incident with u_j is $S_j = \sum_{k=1}^{n} X_{j,k}$ and $\mathbb{E}_{\mathcal{G}_p}(S_j) = np$. We can now apply Chernoff's inequality (Theorem 3.1), letting $k = n^{1/2+\varepsilon/3}$. Then for large enough n

$$\mathbb{P}_{\mathcal{G}_p}\big(S_j \leq np - n^{1/2+\varepsilon/3}\big) \leq e^{-n^{1+2\varepsilon}/2np} = e^{-n^{2\varepsilon}/2p} \leq \tfrac{1}{2}e^{-n^{3\varepsilon/2}} \tag{4.2}$$

and

$$\mathbb{P}_{\mathcal{G}_p}\big(S_j \geq np + n^{1/2+\varepsilon/3}\big) \leq e^{\frac{-n^{1+2\varepsilon}}{2(np+n^{1/2+\varepsilon}/3)}} \leq \tfrac{1}{2}e^{-n^{3\varepsilon/2}}, \tag{4.3}$$

where the last steps on lines (4.2) and (4.3) use $p < 1$. Hence,

$$\mathbb{P}_{\mathcal{G}_p}\big(|S_j - np| \geq n^{1/2+\varepsilon/3}\big) \leq e^{-n^{3\varepsilon/2}}.$$

□

Corollary 4.2. *Fix $0 < a < \frac{1}{2}$. Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.6 and let λ be the random variable which returns the edge density. Then as $m, n \to \infty$ subject to Conditions 4.1,*

$$\mathbb{P}_{\mathcal{G}_p}\left(|\lambda - p| \geq n^{-1+3\varepsilon}\right) \leq e^{-n^{2\varepsilon}}$$

Proof. Simply note that $n^{-1}|S - np| = |\lambda - p|$ so this is actually is an equivalent statement to Lemma 4.1a. □

Lemma 4.3. *Fix $0 < a < \frac{1}{2}$. Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.6.*

1. *Fix any* white *vertex u_j. Let S_j be the random variable that returns the degree of u_j. Then as $m, n \to \infty$ subject to Conditions 4.1,*

$$\mathbb{P}_{\mathcal{G}_p}\left(|S_j - S| \geq n^{1/2+2\varepsilon/5}\right) \leq e^{-n^{4\varepsilon/3}}$$

2. *Fix any* black *vertex v_k. Let T_k be the random variable that returns the degree of v_k. Then as $m, n \to \infty$ subject to Conditions 4.1,*

$$\mathbb{P}_{\mathcal{G}_p}\left(|T_k - T| \geq n^{1/2+10\varepsilon/11}\right) \leq e^{-n^{5\varepsilon/4}}$$

Proof. (of 1.)
Let us fix $1 \leq j \leq m$ and consider S_j. We have the following bounds on $|S_j - np|$ and $|s - np|$ by parts a and b of Lemma 4.1 respectively.

$$\mathbb{P}_{\mathcal{G}_p}\left(|S - np| \geq n^{1/2+\varepsilon/3}\right) \leq e^{-n^{3\varepsilon/2}}$$

$$\mathbb{P}_{\mathcal{G}_p}\left(|S_j - np| \geq n^{1/2+\varepsilon/3}\right) \leq e^{-n^{3\varepsilon/2}}$$

Hence by the triangle inequality we conclude that for large enough n,

$$\mathbb{P}_{\mathcal{G}_p}\left(|S_j - S| \geq n^{1/2+2\varepsilon/5}\right) \leq e^{-n^{4\varepsilon/3}},$$

as required. □

Proof. (of 2.) The symmetry of our bipartite graph allows us to swap the *black* vertices of our random graph for the *white* vertices. Hence by symmetry, any true statement on the variables m, n, u_j, S_j, S implies the corresponding statement on the variables m, n, v_k, T_k, T.

Therefore our bounds on $|S_j - S|$ in Lemma 4.1 (1) imply by symmetry,

$$\mathbb{P}_{\mathcal{G}_p}\left(|T_k - T| \geq m^{1/2+10\varepsilon/11}\right) \leq e^{-m^{4\varepsilon/3}}. \tag{4.4}$$

As we know that $m = o(n^{1+\varepsilon})$ and $n = o(m^{1+\varepsilon})$ the bound above translates into a bound in terms of n. In the working below c, c' are arbitrary constants in $\mathbb{R}$.

$$m^{1/2+2\varepsilon/5} \leq (cn^{1+\varepsilon})^{1/2+2\varepsilon/5} = c'n^{1/2+2\varepsilon/5+\varepsilon/2+2\varepsilon^2/5} \leq n^{1/2+10\varepsilon/11}$$

$$e^{-m^{4\varepsilon/3}} \leq e^{-(cn^{\frac{1}{1+\varepsilon}})^{4\varepsilon/3}} = e^{-n^{4\varepsilon(1-\varepsilon+\varepsilon^2-\varepsilon^3\ldots)/3}c^{4\varepsilon/3}} \leq e^{-n^{5\varepsilon/4}}$$

Hence (4.4) implies:

$$\mathbb{P}_{\mathcal{G}_p}\big(|T_k - T| \geq n^{1/2+10\varepsilon/11}\big) \leq e^{-n^{5\varepsilon/4}}.$$

We have our result. □

The last lemma showed that the degree of a particular vertex on either side is unlikely to vary too greatly from the average of the degrees on that side. This is an important step in showing that a degree sequence $(\boldsymbol{s}, \boldsymbol{t})$ for a random graph in the graph p-model, $\mathcal{G}_p$ is highly likely to be (ε, a)-*regular*.

Lemma 4.4. *Fix $0 < a < \frac{1}{2}$ and $\varepsilon > 0$. Let $\mathcal{G}_p = \mathcal{G}_p(m, n)$ be as in Definition 1.6. Let $S_1, \ldots, S_m$ and $T_1, \ldots, T_n$ be the random variables that return the* degrees *of the* white *vertices $u_1, \ldots, u_m$ and the* black *vertices $v_1, \ldots, v_n$ respectively.*

Then as $m, n \to \infty$ subject to Conditions 4.1,

$$\mathbb{P}_{\mathcal{G}_p}\big(\forall j, k, \quad T_k - T, \;\; S_j - S \text{ uniformly } o(n^{1/2+\varepsilon})\big) \geq 1 - e^{-n^{6\varepsilon/5}}.$$

Proof. By Lemma 4.3, parts (1) and (2) we have the following results:

$$\mathbb{P}_{\mathcal{G}_p}\big(|S_j - S| \geq n^{1/2+2\varepsilon/5}\big) \leq e^{-n^{4\varepsilon/3}},$$
$$\mathbb{P}_{\mathcal{G}_p}\big(|T_k - T| \geq n^{1/2+10\varepsilon/11}\big) \leq e^{-n^{5\varepsilon/4}}.$$

Hence,

$$\begin{aligned}\mathbb{P}_{\mathcal{G}_p}\big(\forall j, \quad |S_j - S| < n^{1/2+2\varepsilon/5} \text{ and } \forall k, \quad |T_k - T| < n^{1/2+10\varepsilon/11}\big) &\geq 1 - me^{-n^{4\varepsilon/3}} - ne^{-n^{5\varepsilon/4}} \\ &\geq 1 - e^{-n^{6\varepsilon/5}}.\end{aligned}$$

□

4.1.2 Edge density in $\mathcal{G}_p$

Recall the term *edge density*, denoted λ and defined by $\lambda = \frac{1}{mn}\sum_j s_j$, i.e. the number of edges divided by total number of possible edges in the complete bipartite graph. There is a condition on λ in Definition 2.7 of an (ε, a)-*regular* degree sequence. We show that this condition is likely to hold in $\mathcal{G}_p$ for suitable values of p. In particular, we show that when p is *acceptable* for a, m and n it is highly likely that λ is *acceptable* for $a+\varepsilon, m$ and n.

First we prove two technical lemmas. The results are a property of the relationship between p and λ where p is a parameter of $\mathcal{G}_p$ and the random variable, λ, returns the edge density of a random graph in $\mathcal{G}_p$.
We prove a lemma showing the bounds placed on $r(1-r)$ by the assumption that r is *acceptable* for a, m and n.

Lemma 4.5. *Let n and m be positive integers. Suppose there exists some $0 < a < \frac{1}{2}$, such that r is* acceptable *for a, m and n.*

Then for $n > e^{16}$,

$$\frac{1}{\ln n} < r(1-r)$$

Proof. By the definition of *acceptable* for a, m and n we know r satisfies:

$$\frac{(1-2r)^2}{4r(1-r)}\Big(1 + \frac{5m}{6n} + \frac{5n}{6m}\Big) \leq a \ln n$$

We observe,

$$\frac{(1-2r)^2}{4r(1-r)} = \frac{1}{4r(1-r)} - 1 \qquad \text{and} \qquad 1 + \frac{5m}{6n} + \frac{5n}{6m} \geq \frac{8}{3}.$$

Hence,

$$\begin{aligned} r(1-r) &> \frac{4}{16 + 3\ln n} \\ &> \frac{1}{\ln n} \qquad\qquad \text{for } n > e^{16} \end{aligned}$$

as required.

□

Lemma 4.6. *Fix $\varepsilon > 0$. Let $\mathcal{G}_p = \mathcal{G}_p(m, n)$ be as in Definition 1.6. Then as $m, n \to \infty$ subject to Conditions 4.1,*

$$\mathbb{P}_{\mathcal{G}_p}\Big(\frac{p(1-p)}{\lambda(1-\lambda)} = 1 + O(n^{-1+4\varepsilon})\Big) \geq 1 - e^{-n^{2\varepsilon}}.$$

Proof. We begin by noting the following rearrangement,

$$\frac{p(1-p)}{\lambda(1-\lambda)} = 1 - (p-\lambda)\Big(\frac{\lambda - (1-p)}{p(1-p)}\Big)^{-1}. \tag{4.5}$$

By Lemma 4.5, because p is *acceptable* for a, m and n:

$$p(1-p) > \tfrac{1}{\ln n}. \tag{4.6}$$

Also note that $|\lambda + p - 1| < 1$. Thus by (4.6),

$$\Big|\frac{\lambda - (1-p)}{p(1-p)}\Big| < \ln n. \tag{4.7}$$

The expected value in the graph p-model, $\mathcal{G}_p$ of the edge-density, λ, is p. So, when λ is equal to its expectation, (4.5) is precisely 1. An earlier result, corollary 4.2 provides the following bound:

$$\mathbb{P}_{\mathcal{G}_p}\big(|\lambda - p| \geq n^{-1+3\varepsilon}\big) \leq e^{-n^{2\varepsilon}}.$$

Hence by (4.5), (4.7) and noting that $\ln n < n^{\varepsilon}$ for large enough n, we have the result.

□

We are now in a position to prove our result on *acceptable* degree sequences.

Lemma 4.7. *Fix $\varepsilon > 0$ and $0 < a < \frac{1}{2}$. Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.6. Then as $m, n \to \infty$ subject to Conditions 4.1,*

$$\mathbb{P}_{\mathcal{G}_p}\Big(\lambda \textit{ is } \text{acceptable} \textit{ for } a+\varepsilon, m \textit{ and } n\Big) \geq 1 - e^{-n^{3\varepsilon/2}}.$$

Proof. For convenience we let $c = \Big(1 + \frac{5m}{6n} + \frac{5n}{6m}\Big)^{-1}$. Note that $c > n^{-\varepsilon}$. The condition that p is *acceptable* for a, m and n can now be written

$$\frac{1}{4pq} < 1 + ac\ln n. \tag{4.8}$$

By Lemma 4.6,

$$\mathbb{P}_{\mathcal{G}_p}\Big(\frac{p(1-p)}{\lambda(1-\lambda)} = 1 + O(n^{-1+4\varepsilon})\Big) \geq 1 - e^{-n^{2\varepsilon}}.$$

We work now in the graph p-model, $\mathcal{G}_p$; with probability greater than $1 - e^{-n^{3\varepsilon/2}}$ the following holds.

$$
\begin{aligned}
\frac{1}{4\lambda(1-\lambda)} &= \frac{1}{4pq}\big(1 + O(n^{-1+4\varepsilon})\big) && (4.9)\\
&\leq \frac{1}{4pq} + O(n^{-1+4\varepsilon}\ln n) && (4.10)\\
&< 1 + ac\ln n + O(n^{-1+4\varepsilon}\ln n) && (4.11)\\
&< 1 + \Big(a + O\Big(\frac{n^{-1+4\varepsilon}\ln n}{c\ln n}\Big)\Big)c\ln n && (4.12)\\
&< 1 + \Big(a + O\big(n^{-1+5\varepsilon}\big)\Big)c\ln n && (4.13)\\
&< 1 + \big(a + \varepsilon\big)c\ln n && (4.14)
\end{aligned}
$$

We justify the steps line by line. Line (4.9) holds with probability at least $1 - e^{-n^{2\varepsilon}}$ by Lemma 4.6. Note this is the one step that holds only probabilistically. By Lemma 4.5 because p is *acceptable* for a, m and n, we know that $\frac{1}{pq} > \ln n$. This implies (4.10). Line (4.11) now follows by (4.8). Then (4.12) is a rearrangement of (4.11). Note that $c > n^{-\varepsilon}$ because of our conditions on m and n. Thus we have (4.13). We are considering n tending to infinity so ε is larger than $O(n^{-1/2+3\varepsilon/2})$. Hence (4.14), which is exactly as we require. □

4.1.3 Bounding result on (ε, a)-*regular* degree sequences

When our parameter p is *acceptable* for a, m and n there is a high probability that a graph chosen randomly in $\mathcal{G}_p$ will have an (ε, a)-*regular* degree sequence. We state and prove this formally in the following lemma.

Lemma 4.8. *Fix $0 < a < \frac{1}{2}$ and $\varepsilon > 0$ such that $a' = a + \varepsilon < \frac{1}{2}$. Let $\mathcal{G}_p = \mathcal{G}_p(m, n)$ be as in Definition 1.6. Then as $m, n \to \infty$ subject to Conditions 4.1,*

$$\mathbb{P}_{\mathcal{G}_p}\big((\boldsymbol{S}, \boldsymbol{T}) \text{ is } (\varepsilon, a)\text{-regular}\big) \geq 1 - e^{-n^{7\varepsilon/6}}$$

Proof. By Lemmas 4.4 and 4.7. Note that $e^{-n^{7\varepsilon/6}} < e^{-n^{3\varepsilon/2}} + e^{-n^{6\varepsilon/5}}$ for large enough n. The result now follows. □

Recall that this is important as it is only graphs with (ε, a)-*regular* degree sequences $(\boldsymbol{s}, \boldsymbol{t})$ which satisfy Greenhill and McKay's enumeration result, Theorem 2.8. Hence Lemma 4.8 shows that for a degree sequence $(\boldsymbol{s}, \boldsymbol{t})$, generated at random in the p-model, $\mathcal{G}_p$, there is a high probability that we can apply Theorem 2.8.

In the next section we will show that in the graph p-model, $\mathcal{G}_p$, with a high probability, a random graph will have a degree sequence $(\boldsymbol{s}, \boldsymbol{t})$ that is non-*pathological*.

4.2 *Pathological* degree sequences in $\mathcal{G}_p$.

We show that in the graph p-model, $\mathcal{G}_p$, the probability of selecting a graph whose degree sequence is *pathological* is very small. This is done in Theorem 4.9 at the end of this section. Recall that a degree sequence is *pathological* if it is either not (ε, a)-*regular* or is such that the expression (3.5) is too large. Earlier in Section 4.1, we found that the first of these possibilities was unlikely, i.e. we showed that the probability of a degree sequence not being (ε, a)-*regular* was small. In this section we concentrate on the other property that leads to $(a, b, m, n, \varepsilon)$-*pathological* degree sequences. Thus, we analyse the left hand side of (3.5). In particular, we want to bound the probability in the graph p-model, $\mathcal{G}_p$, that the left hand side of (3.5) exceeds $O(n^{-b})$. We briefly outline the steps we take to achieve this.

By symmetry we will see it is sufficient to show that $1 - \frac{\sum_j (S_j - S)^2}{\lambda(1-\lambda)mn}$ is $O(n^{-b})$ with high probability in $\mathcal{G}_p$. This we do in two steps, we show that both of the following hold with high probability in $\mathcal{G}_p$.

$$\frac{\sum_j (S_j - S)^2 - pqmn}{\lambda(1-\lambda)mn} = O(n^{-b}) \tag{4.15}$$

$$\frac{pq}{\lambda(1-\lambda)} = \frac{pqmn}{\lambda(1-\lambda)mn} = 1 + O(n^{-b}) \tag{4.16}$$

Proving (4.15) comprises the entirety of Sections 4.2.1 and 4.2.2. It is a concentration result, we show that the function on the *white* degrees $\sum_j (S_j - S)^2$ is likely to be close to $pqmn$. In Section 4.2.1 we define a new random variable, S_j^*, effectively allowing us to work only with degree sequences whose *white* degrees are close to their mean. Then, in Section 4.2.2, working with these new random variables we use concentration inequalities to prove (4.15) holds with high probability in $\mathcal{G}_p$.

The result (4.16) follows more readily; indeed we have already proven a stronger result, Lemma 4.6. It was proven using bounds on $\lambda(1 - \lambda)$ and $|p - \lambda|$.

Finally, in Section 4.2.3, the two results (4.15) and (4.16) are combined with the result of Lemma 4.8, which bounds the probability that a degree sequence will not be (ε, a)-*regular* in the graph p-model, $\mathcal{G}_p$. The combination of these yields Theorem 4.9 which states that the probability of a degree sequence being *pathological* is asymptotically very small in the graph p-model, $\mathcal{G}_p$.

4.2.1 Restricted distribution on the degrees of vertices

We will introduce a truncated version of our random variables for the degrees of the *white* vertices.

Definition 4.1 (truncated degree, S_j^*)**.**

$$S_j^* = \max\big(np - n^{1/2+\varepsilon}, \min(np + n^{1/2+\varepsilon}, S_j)\big)$$

We will write $\boldsymbol{S}^* := (S_1^*, \ldots, S_m^*)$. The random variable S_j^* has a truncated distribution of S_j. Note

$$\mathbb{P}_{\mathcal{G}_p}\big(S_j^* = x\big) \quad := \quad \begin{cases} \mathbb{P}_{\mathcal{G}_p}\big(S_j = x\big) & \text{if } |x - np| < n^{1/2+\varepsilon} \\ \mathbb{P}_{\mathcal{G}_p}\big(S_j \geq np + n^{1/2+\varepsilon}\big) & \text{if } x = np + n^{1/2+\varepsilon} \\ \mathbb{P}_{\mathcal{G}_p}\big(S_j \leq np - n^{1/2+\varepsilon}\big) & \text{if } x = np - n^{1/2+\varepsilon} \\ 0 & \text{otherwise} \end{cases}$$

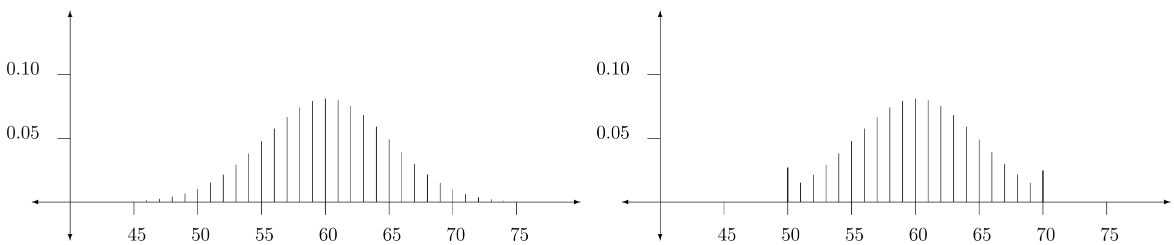


Figure 4.1: Probability distributions of s_1 and s_1^* for $n = 100$, $p = 0.6$ and $\varepsilon = 0$.

The advantage of our new random variables is twofold: each S_j^* is at most $\sqrt{n}$ away from the expected mean of the degrees of the *white* vertices, np, but also agrees with our original random variable S_j on a large part of the domain. We make this claim of *likely agreement* precise in the following lemma.

Lemma 4.2. *Define* $D_p = (np - n^{1/2+\varepsilon}, np + n^{1/2+\varepsilon})^m \subset \mathbb{R}^m$.
Then,

$$\mathbb{P}_{\mathcal{G}_p}\big(\boldsymbol{S} \in D_p\big) \geq 1 - e^{-n^{4\varepsilon/3}} \tag{4.17}$$

and for any m-tuple $\boldsymbol{x} \in D_p$,

$$\mathbb{P}_{\mathcal{G}_p}\big(\boldsymbol{S}^* = \boldsymbol{x}\big) = \mathbb{P}_{\mathcal{G}_p}\big(\boldsymbol{S} = \boldsymbol{x}\big). \tag{4.18}$$

Proof. Fix j. Note that by the definition of our new random variable S_j^*, we have that for $y \in (np - n^{1/2+\varepsilon}, np + n^{1/2+\varepsilon})$,

$$\mathbb{P}_{\mathcal{G}_p}\big(S_j^* = y\big) = \mathbb{P}_{\mathcal{G}_p}\big(S_j = y\big).$$

The extension to the m-dimensional probability space is then clear so we have (4.18). To prove (4.17) we recall that degree of each *white* vertex is highly likely to concentrated about its mean, np. By Lemma 4.1,

$$\mathbb{P}_{\mathcal{G}_p}\big(|S_j - np| \geq n^{1/2+\varepsilon/3}\big) \leq e^{-n^{3\varepsilon/2}}$$

hence,

$$\mathbb{P}_{\mathcal{G}_p}\big(\forall j, |S_j - np| \leq n^{1/2+\varepsilon/3}\big) \geq 1 - ne^{-n^{3\varepsilon/2}} \geq 1 - e^{-n^{4\varepsilon/3}}. \tag{4.19}$$

The result (4.17) follows directly from equation (4.19) and so we are done.

□

We now work with these new random variables S_j^*, where the distribution of the degree of each vertex is restricted.

4.2.2 Bounding a function on the *white* degrees: $\sum_j (S_j - S)^2$.

The aim of this subsection is to show that the absolute difference between the expressions $\sum_j (S_j - S)^2$ and $pqmn$ is asymptotically likely to be small in the graph p-model, $\mathcal{G}_p$. This is done in Lemma 4.6 at the end of this subsection.

We begin with some algebraic trickery and note:

$$\sum_{j=1}^{m} (S_j - S)^2 = -m(S - np)^2 + \sum_{j=1}^{m} (S_j - np)^2. \tag{4.20}$$

Line (4.20) and the triangle inequality now allow us to write:

$$\begin{aligned} \big|\sum_j (S_j - S)^2 - pqmn\big| \leq\ & |m(S-np)^2| + \big|\sum_j (S_j - np)^2 - \sum_j (S_j^* - np)^2\big| \\ & + \big|\sum_j (S_j^* - np)^2 - \mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j^* - np)^2\big)\big| + \big|\mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j^* - np)^2\big) - pqmn\big|. \end{aligned} \tag{4.21}$$

We note that when our m-tuple , $\boldsymbol{S}$, for the degrees of the *white* vertices, lies within the restriction of all m-tuples, $\boldsymbol{S} \in D_p$ the following equality holds,

$$\sum_j (S_j - np)^2 = \sum_j (S_j^* - np)^2.$$

In particular, whenever $\boldsymbol{S} \in D_p$, then by (4.21), we have

$$\begin{aligned} \big|\sum_j (S_j - S)^2 - pqmn\big| \leq\ & |m(S-np)^2| + \big|\sum_j (S_j^* - np)^2 - \mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j^* - np)^2\big)\big| \\ & + \big|\mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j^* - np)^2\big) - pqmn\big|. \end{aligned} \tag{4.22}$$

As we shall see next, each of three terms on the right hand side of (4.22) is asymptotically very likely to be small in the graph p-model, $\mathcal{G}_p$. We prove probabilistic bounds for these three terms in Lemmas 4.3, 4.4 and 4.5 respectively. The first bound follows directly from a bound shown in the previous section. To achieve the middle bound we use a concentration inequality while the last term can be bounded by properties of expectation and of our random variables S_j and S_j^*.

We begin by bounding the first of these three terms on the right-hand side of (4.22) and show that the magnitude of $m(S-np)^2$ is small with high probability in $\mathcal{G}_p$.

Lemma 4.3. *Fix $0 < a < \frac{1}{2}$ and $\varepsilon > 0$. Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.6. Let S be the random variable that returns the average degree of the* white *vertices $u_1, \dots, u_m$. Then as $m, n \to \infty$ subject to Conditions 4.1,*

$$\mathbb{P}_{\mathcal{G}_p}\big(|m(S-np)^2| \geq n^{1+8\varepsilon}\big) \leq e^{-n^{2\varepsilon}}$$

Proof. This follows by a previous result. By Lemma 4.1 (1), $\mathbb{P}_{\mathcal{G}_p}\big(|S-np| \geq n^{3\varepsilon}\big) \leq e^{-n^{2\varepsilon}}$. An immediate corollary of this is

$$\mathbb{P}_{\mathcal{G}_p}\big(m(S-np)^2 \geq m(n^{3\varepsilon})^2\big) \leq e^{-n^{2\varepsilon}}.$$

Then, since $n^{1+8\varepsilon} \geq m(n^{3\varepsilon})^2$ we have our result. □

Lemma 4.4. *Fix $0 < a < \frac{1}{2}$ and $\varepsilon > 0$. Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.6. For $1 \leq j \leq m$, let S_j^* be the random variable which returns the* truncated degree *of the* white *vertex u_j. Then as $m, n \to \infty$ subject to Conditions 4.1,*

$$\mathbb{P}_{\mathcal{G}_p}\big(|\sum_j (S_j^*-np)^2 - \mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j^*-np)^2\big)| \geq n^{3/2+4\varepsilon}\big) \leq e^{-n^{2\varepsilon}}.$$

Proof. Fix a *white* vertex u_j and consider its *truncated* degree S_j^*. By Definition 4.1 (*truncated* degree) : $|S_j^*-np| \leq n^{1/2+\varepsilon}$. This implies $(S_j^*-np)^2 \leq n^{1+2\varepsilon}$. Because the last result is true always, we can say the same for the expectation, i.e. that $\mathbb{E}_{\mathcal{G}_p}\big((S_j^*-np)^2\big) \leq n^{1+2\varepsilon}$. Since both $(S_j^*-np)^2$ and its expectation, $\mathbb{E}_{\mathcal{G}_p}\big((S_j^*-np)^2\big)$, are positive, we have

$$|(S_j^*-np)^2 - \mathbb{E}_{\mathcal{G}_p}\big((S_j^*-np)^2\big)| \leq |\max\{(S_j^*-np)^2, \mathbb{E}_{\mathcal{G}_p}\big((S_j^*-np)^2\big)\}| \leq n^{1+2\varepsilon}. \quad (4.23)$$

These observations above hold for all $1 \leq j \leq m$. Define $X_j = (S_j^*-np)^2$. Then because the degrees (and hence the *truncated* degrees) of the *white* vertices are independent, $X_1, \dots, X_n$ are independent random variables. By (4.23), these random variables satisfy $|X_j - E(X_j)| \leq n^{1+2\varepsilon}$ for $1 \leq j \leq m$. Hence we can apply Theorem 3.2 to get

$$\mathbb{P}_{\mathcal{G}_p}\big(|\sum_j (S_j^*-np)^2 - \mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j^*-np)^2\big)| \geq k\big) \leq e^{\frac{-k^2}{2m(n^{1+2\varepsilon})^2}} \leq e^{\frac{-k^2}{n^{\varepsilon}n^{1+\varepsilon}n^{2+4\varepsilon}}} \leq e^{\frac{-k^2}{n^{3+6\varepsilon}}}.$$

Substituting $k = n^{3/2+4\varepsilon}$ gives the required result.

□

Lemma 4.5. *Fix $\varepsilon > 0$. Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.6. For $1 \le j \le m$ let S_j^* be the random variable which returns the* truncated degree *of the* white *vertex u_j. Then as $m,n \to \infty$ subject to $m = o(n^{1+\varepsilon})$ and $n = o(n^{1+\varepsilon})$,*

$$|\mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j - np)^2\big) - pqmn| \;\le\; e^{-n^{7\varepsilon/6}}.$$

Proof. By calculations done in appendix, line (8.4), we know that

$$\mathbb{E}_{\mathcal{G}_p}\big(\textstyle\sum_j (S_j - np)^2\big) = pqmn.$$

Hence,

$$|\mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j^* - np)^2\big) - pqmn| = |\mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j^* - np)^2\big) - \mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j - np)^2\big)|. \tag{4.24}$$

So it is equivalent to bound the right-hand-side of (4.24). The proof now follows by properties of expectation. Firstly, by the definition of expectation

$$\begin{aligned}
\mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j - np)^2\big) &= \sum_{\boldsymbol{x}} \Big[\mathbb{P}_{\mathcal{G}_p}\big((S_1,\dots,S_m) = \boldsymbol{x}\big) \sum_j (x_j - np)^2\Big] \\
\mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j^* - np)^2\big) &= \sum_{\boldsymbol{x}} \Big[\mathbb{P}_{\mathcal{G}_p}\big((S_1^*,\dots,S_m^*) = \boldsymbol{x}\big) \sum_j (x_j - np)^2\Big].
\end{aligned}$$

For any $\boldsymbol{x} \in D_p$, i.e. most degree sequences, $\boldsymbol{x}$, of the *white* vertices,

$$\mathbb{P}_{\mathcal{G}_p}\big(\boldsymbol{S^*} = \boldsymbol{x}\big) \;=\; \mathbb{P}_{\mathcal{G}_p}\big(\boldsymbol{S} = \boldsymbol{x}\big).$$

Hence for $\boldsymbol{x} \in D_p$ the corresponding terms in the sum below will cancel. Thus,

$$\begin{aligned}
|\mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j - np)^2\big) - \mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j^* - np)^2\big)| &\le \mathbb{P}_{\mathcal{G}_p}\big((\boldsymbol{s},\boldsymbol{t}) : \boldsymbol{s} \notin D_p\big) \max_{\boldsymbol{x}} \sum_j (x_j - np)^2 \\
&\le e^{-n^{6\varepsilon/5}} mn^2 \\
&\le e^{-n^{7\varepsilon/6}}
\end{aligned}$$

as required. □

These three results (Lemmas 4.3,4.4 and 4.5) allow us to prove (4.15); indeed, we show a slightly stronger result in the following lemma.

Lemma 4.6. *Fix $0 < a < \frac{1}{2}$ and $\varepsilon > 0$ such that $a = a' + \varepsilon < \frac{1}{2}$. Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.6.*

For $1 \le j \le m$, let S_j be random variable that returns the degree *of the* white *vertex u_j, let S return the average of these degrees and let λ return the edge density.*

Then as $m, n \to \infty$ subject to Conditions 4.1,

$$\mathbb{P}_{\mathcal{G}_p}\left(\Big| \frac{\sum_j (S_j - S)^2 - pqmn}{\lambda(1-\lambda)mn} \Big| \ \ge\ n^{-1/2+8\varepsilon} \right) \ \le\ e^{-n^{6\varepsilon/5}}$$

Proof. In the Lemmas 4.4, 4.3 and 4.5 we compiled the following probability bounds:

$$\mathbb{P}_{\mathcal{G}_p}\Big(|m(S-np)^2| \ \ge\ n^{3/2+5\varepsilon} \Big) \le e^{-n^{1/2-3\varepsilon}} \tag{4.25}$$

$$\mathbb{P}_{\mathcal{G}_p}\Big(|\sum_j (S_j - np)^2 - \mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j - np)^2\big)| \ \ge\ n^{3/2+3\varepsilon} \Big) \le e^{-n^{3\varepsilon/2}} \tag{4.26}$$

$$\Big| \mathbb{E}_{\mathcal{G}_p}\big(\sum_j (S_j^* - np)^2\big) - pqmn \Big| \ \le\ e^{-n^{1/2}}. \tag{4.27}$$

By (4.22), for any $\boldsymbol{s} \in D_p$, our expression of interest, $\big|\sum_j (s_j - s)^2 - pqmn\big|$, is bounded above by the sum of the three terms in (4.25), (4.26) and (4.27).

Hence,

$$\mathbb{P}_{\mathcal{G}_p}\Big(|\sum_j (S_j - S)^2 - pqmn| \le n^{3/2+5\varepsilon} + n^{3/2+3\varepsilon} + e^{-n^{1/2}} \Big) \ge 1 - e^{-n^{1/2-3\varepsilon}} - e^{-n^{3\varepsilon/2}} - \mathbb{P}_{\mathcal{G}_p}\big(\boldsymbol{s} \notin D_p\big)$$

By Lemma 4.2, $\mathbb{P}_{\mathcal{G}_p}\big(\boldsymbol{s} \notin D_p\big) \le e^{-n^{6\varepsilon/5}}$. Also, $1 - e^{-n^{1/2-3\varepsilon}} - e^{-n^{3\varepsilon/2}} - e^{-n^{6\varepsilon/5}} \ge 1 - e^{-n^{7\varepsilon/6}}$. So we have

$$\mathbb{P}_{\mathcal{G}_p}\Big(|\sum_j (S_j - S)^2 - pqmn| \le n^{3/2+6\varepsilon} \Big) \ge 1 - e^{-n^{7\varepsilon/6}}.$$

By Lemma 4.7, λ is *acceptable* for a, m and n with probability at least $1 - e^{-n^{3\varepsilon/2}}$. Hence, by Lemma 4.5, $\lambda(1-\lambda) > \frac{1}{\ln n} > n^{-\varepsilon/2}$, with this same probability. Hence,

$$\mathbb{P}_{\mathcal{G}_p}\Big(\frac{|\sum_j (S_j - S)^2 - pqmn|}{\lambda(1-\lambda)mn} \le n^{-1/2+9\varepsilon}\Big) \ge 1 - e^{-n^{8\varepsilon/7}}. \quad \square$$

This Lemma 4.6 together with Lemma 4.6 will enable us to show that for a random graph in $\mathcal{G}_p$ its degree sequence $(\boldsymbol{s}, \boldsymbol{t})$, is unlikely to be *pathological*.

Corollary 4.7. *Fix $0 < a < \frac{1}{2}$ and $\varepsilon > 0$ such that $a := a' + \varepsilon < \frac{1}{2}$. Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.6.*
For $1 \le j \le m$, let S_j be random variable that returns the degree *of the* white *vertex u_j, let S return the average of these degrees and let λ return the edge density.*

Then as $m, n \to \infty$ subject to Conditions 4.1,

$$\mathbb{P}_{\mathcal{G}_p}\left(\left|1 - \frac{\sum_j (S_j - S)^2}{\lambda(1-\lambda)mn}\right| \;\ge\; n^{-1/2+9\varepsilon} \right) \;\le\; e^{-n^{7\varepsilon/6}}$$

Proof. This follows directly by Lemmas 4.6 and 4.6. $\square$

In the next section we will use this result, Corollary 4.7 and some symmetry arguments to show that *pathological* degree sequences are rare in the graph p-model, $\mathcal{G}_p$.

4.2.3 Likelihood of *pathological* degree sequences

Now that we have shown Corollary 4.7 we show that a similar probabilistic bound hold for the degrees of the *black* vertices.

Lemma 4.8. *Fix $0 < a < \frac{1}{2}$ and let $\varepsilon > 0$ such that $a' := a + \varepsilon < \frac{1}{2}$. Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.6.*
For $1 \le k \le n$, let T_k be random variable that returns the degree *of the* black *vertex v_k, let T return the average of these degrees and let λ return the edge density.*

Then as $m, n \to \infty$ subject to Conditions 4.1,

$$\mathbb{P}_{\mathcal{G}_p}\left(\left|1 - \frac{\sum_k (T_k - T)^2}{\lambda(1-\lambda)mn}\right| \;\ge\; n^{-1/2+9\varepsilon} \right) \;\le\; e^{-n^{8\varepsilon/7}}.$$

Proof. This proof will be done utilising symmetry.

Note that because $n = o(m^{1+\varepsilon})$, $\ln n < (1+\varepsilon)\ln m$. Hence,

$$a \ln n < (a' - \varepsilon)(1+\varepsilon)\ln m < a' \ln m \tag{4.28}$$

Now, because p is *acceptable* for a, m and n and by (4.28) we can conclude that p is *acceptable* for a', n and m. Then by Corollary 4.7 swapping each instance of 'S' for 'T' and each instance of 'm' for 'n'. We can conclude that as $m, n \to \infty$ the following holds.

$$\mathbb{P}_{\mathcal{G}_p}\left(\left| 1 - \frac{\sum_k (T_k - T)^2}{\lambda(1-\lambda)mn} \right| \;\geq\; m^{-1/2+8\varepsilon} \right) \;\leq\; e^{-m^{6\varepsilon/5}}. \tag{4.29}$$

Now the inequalities,

$$m^{1/2+8\varepsilon} \leq (cn^{1+\varepsilon})^{1/2+8\varepsilon} = c'n^{1/2+8\varepsilon+\varepsilon/2+8\varepsilon^2} \leq n^{1/2+9\varepsilon}$$
$$e^{-m^{7\varepsilon/6}} \leq e^{-(cn^{\frac{1}{1+\varepsilon}})^{7\varepsilon/6}} = e^{-n^{7\varepsilon(1-\varepsilon+\varepsilon^2-\varepsilon^3\ldots)/6}c^{7\varepsilon/6}} \leq e^{-n^{8\varepsilon/7}}$$

complete the proof. □

Theorem 4.9. *Fix $a, b \in \mathbb{R}^+$ and $\varepsilon > 0$ such that $a+b < \frac{1}{2}$, $a' = a+\varepsilon < \frac{1}{2}$ and $b+17\varepsilon < 1$. Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition. Then as $m, n \to \infty$ subject to Conditions 4.1,*

$$\mathbb{P}_{\mathcal{G}_p}\big(\; (\boldsymbol{S}, \boldsymbol{T}) \;\text{ is } (a,b,m,n,\varepsilon)\text{-}\textit{pathological}\; \big) \leq e^{-n^{10\varepsilon/11}}.$$

Proof. Together, Corollary 4.7 and Lemma 4.8 imply,

$$\mathbb{P}_{\mathcal{G}_p}\left(\left| \left(1 - \frac{\sum_j (S_j - S)^2}{\lambda(1-\lambda)mn}\right)\left(1 - \frac{\sum_k (T_k - T)^2}{\lambda(1-\lambda)mn}\right) \right| \;\geq\; n^{-1+17\varepsilon} \right) \;\leq\; e^{-n^{9\varepsilon/10}} \tag{4.30}$$

Note that $n^{-1+17\varepsilon} = O(n^{-b})$ for $b < 1 - 17\varepsilon$. We are almost done. To finish, recall by Lemma 4.8,

$$\mathbb{P}_{\mathcal{G}_p}\big((\boldsymbol{S}, \boldsymbol{T}) \text{ is not } (\varepsilon, a)\text{-}\textit{regular}\big) \leq e^{-n^{7\varepsilon/6}}.$$

□

Chapter 5

Graph *edge*-model, $\mathcal{G}_M$.

This chapter concerns the graph *edge*-model, $\mathcal{G}_M$ (see Definition 1.7 on page 5). In this model every bipartite graph on (m, n) vertices with M edges is equally likely.

We will derive a simple formula for the probability of any non-$(a, b, m, n, \varepsilon)$-*pathological* degree sequence $(\boldsymbol{s}, \boldsymbol{t})$ in the graph *edge*-model, $\mathcal{G}_M$.

The starting point will be the enumeration result by Greenhill and McKay, Theorem 2.8 as in the last chapter on the graph p-model, $\mathcal{G}_p$. Hence the first step will be (again) to show that this formula is likely to be applicable to a random degree sequence in the probability space. That is, we show that a degree sequence in $\mathcal{G}_M$ is very likely to be (ε, a)-*regular*.

Then, to ensure the enumeration result simplifies we will also show that *pathological* degree sequences are rare in $\mathcal{G}_M$. These two results on (ε, a)-*regular* and *pathological* degree sequences are proven in Section 5.2.

To show these two results we will prove similar probabilistic bounds in $\mathcal{G}_M$ to those already shown in the $\mathcal{G}_p$ model. The graph *edge*-model is the restriction of the $\mathcal{G}_p$ model to graphs with M edges. Hence these two probability spaces are related which allows us to derive Lemma 5.1. This lemma provides an upper bound on any event A occurring in $\mathcal{G}_M$ in terms of the probability of the same event A occurring in $\mathcal{G}_p$. Using Lemma 5.1 the results on $\mathcal{G}_M$ in Section 5.2 in this chapter follow from similar results on $\mathcal{G}_p$ in the previous chapter.

We will need to refer to the following conditions.

Conditions 5.1. *These conditions apply to an m-tuple $\boldsymbol{s} = \boldsymbol{s}(m, n) = (s_1, ..., s_m)$ and an n-tuple $\boldsymbol{t} = \boldsymbol{t}(m, n) = (t_1, ..., t_n)$. The conditions depend on the parameters a and ε.*

For $m, n \to \infty$,

- $\frac{M}{mn}$ *is* acceptable *for* a, m *and* n.
- $m = o(n^{1+\varepsilon})$ *and* $n = o(m^{1+\varepsilon})$.

5.1 Relation to graph p-model, $\mathcal{G}_p$.

Lemma 5.1. *Let X be a random variable defined on all bipartite graphs on (m, n) vertices. Then for $p = \frac{M}{mn}$, we have*

$$\mathbb{P}_{\mathcal{G}_M}(X) \leq mn\mathbb{P}_{\mathcal{G}_p}\big(X\big).$$

Proof. The graph *edge*-model, $\mathcal{G}_M$, is the restriction of the graph p-model, $\mathcal{G}_p$, to M edges, so $\mathbb{P}_{\mathcal{G}_M}(X) = \mathbb{P}_{\mathcal{G}_p}\big(X | \#\text{edges } = M\big)$.

Now, let A, B be any random variables. We note by Bayes Theorem (given $\mathbb{P}(B = 1) > 0$):

$$\mathbb{P}\Big(A \mid (B = 1)\Big) = \frac{\mathbb{P}\Big(A \ \& \ (B = 1)\Big)}{\mathbb{P}(B = 1)} \leq \frac{\mathbb{P}(A)}{\mathbb{P}(B = 1)}. \tag{5.1}$$

We work in the graph p-model, $\mathcal{G}_p$, and set the value of our parameter $p = \frac{M}{mn}$.[1] In (5.1) we substitute, for B, the indicator function, I_M for the event that the graph has M edges. Then substitute X for A and we have the following inequality.

$$\mathbb{P}_{\mathcal{G}_M}(X) \leq \frac{\mathbb{P}_{\mathcal{G}_{p=\frac{M}{mn}}}(X)}{\mathbb{P}_{\mathcal{G}_{p=\frac{M}{mn}}}(I_M = 1)}.$$

In the graph p-model, $\mathcal{G}_p$, the number of edges in the graph is between 0 and mn inclusive; indeed the number of edges in the graph is the binomial distribution (mn, p). Here the most likely values for the number of edges are those close to the expected value, pmn. In particular, because we chose $p = \frac{M}{mn}$ then the probability that the number of edges in the graph is M is at least $\frac{1}{mn}$. That is $\mathbb{P}_{\mathcal{G}_{M=\frac{M}{mn}}}(I_M = 1) \geq \frac{1}{mn}$. This yields the desired result. □

5.2 (ε, a)-regular and *pathological* degree sequences in $\mathcal{G}_M$.

Here, we show that in the graph *edge*-model, $\mathcal{G}_M$, (ε, a)-*regular* degree sequences are highly likely.

[1] Note by setting $p = \frac{M}{mn}$, p is *acceptable* for a, m and n if and only if M is *acceptable* for a, m and n.

Lemma 5.1. *Fix $a \in \mathbb{R}^+$ and $\varepsilon > 0$ such that $a < \frac{1}{2}$. Let $\mathcal{G}_M = \mathcal{G}_M(m,n)$ be as in Definition 1.7. Then as $m, n \to \infty$ subject to Conditions 5.1,*

$$\mathbb{P}_{\mathcal{G}_M}(\ (\boldsymbol{S},\boldsymbol{T})\ \text{ is not } (\varepsilon,a)\text{-}\textit{regular}\ ) \leq e^{-n^{8\varepsilon/7}}$$

Proof. By Lemma 4.8,

$$\mathbb{P}_{\mathcal{G}_p}\big((\boldsymbol{S},\boldsymbol{T}) \text{ is } (\varepsilon,a)\text{-}\textit{regular}\big) \geq 1 - e^{-n^{7\varepsilon/6}}$$

Hence we know with probability at least $1 - e^{-n^{6\varepsilon/5}}$, that a degree sequence is likely to be (ε,a)-*regular* in the graph p-model, $\mathcal{G}_p$. So by Lemma 5.1 in the graph *edge*-model, $\mathcal{G}_M$, we know a degree sequence is (ε,a)-*regular* with probability at least $1 - mne^{-n^{7\varepsilon/6}}$, i.e. with probability at least $1 - e^{-n^{8\varepsilon/7}}$.

□

We also bound the probability that a random degree sequence is *pathological* in the graph *edge*-model, $\mathcal{G}_M$.

Theorem 5.2. *Fix $a, b \in \mathbb{R}^+$ and $\varepsilon > 0$ such that $a+b < \frac{1}{2}$, $a' = a+\varepsilon < \frac{1}{2}$ and $b+17\varepsilon < 1$. Let $\mathcal{G}_M = \mathcal{G}_M(m,n)$ be as in Definition 1.7. Then as $m, n \to \infty subject to Conditions 5.1$,*

$$\mathbb{P}_{\mathcal{G}_M}(\ (\boldsymbol{S},\boldsymbol{T})\ \text{ is } (a,b,m,n,\varepsilon)\text{-}\textit{pathological}\ ) \leq e^{-n^{11\varepsilon/12}}.$$

Proof. We proceed directly from Theorem 4.9 in the same manner as the proof of Lemma 5.1.

□

Chapter 6

Graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$.

In our discussions pertaining to the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$, we will often refer to the following set of conditions.

Conditions 6.1. *These conditions apply to an m-tuple $\boldsymbol{s} = \boldsymbol{s}(m,n) = (s_1, ..., s_m)$ and an n-tuple $\boldsymbol{t} = \boldsymbol{t}(m,n) = (t_1, ..., t_n)$. The conditions depend on the parameters a and ε.*

For $m, n \to \infty$,

- *$t_k - t$ is uniformly $O(n^{1/2+\varepsilon})$ for $1 \leq k \leq n$.*
- *$\lambda = \frac{1}{mn}\sum_k t_k$ is* acceptable *for a, m and n.*
- *$m = o(n^{1+\varepsilon})$ and $n = o(m^{1+\varepsilon})$.*

6.1 (ε, a)-regular degree sequences in $\mathcal{G}_{\boldsymbol{t}}$

For a graph, G, chosen at random in $\mathcal{G}_{\boldsymbol{t}}$ we show that with high probability the degree sequence $(\boldsymbol{s}, \boldsymbol{t})$ of G is (ε, a)-*regular*.

Lemma 6.1. *Fix $a \in \mathbb{R}^+$ and $\varepsilon > 0$ such that $a < \frac{1}{2}$. Let $\mathcal{G}_{\boldsymbol{t}} = \mathcal{G}_{\boldsymbol{t}}(m,n)$ be as in Definition 1.8. Then as $m, n \to \infty$ subject to Conditions 6.1,*

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(\ (\boldsymbol{S}, \boldsymbol{T}) \ \text{ is not } (\varepsilon, a)\text{-}\textit{regular}\ ) \leq e^{-n^{6\varepsilon/5}}$$

Proof. By the assumptions on the model, $\mathcal{G}_{\boldsymbol{t}}$, we already have many of the requirements of (ε, a)-*regular*. We have the required condition on the edge density, λ, and also that $t_k - t$ is uniformly $o(n^{1/2+\varepsilon})$ for each graph in the domain of $\mathcal{G}_{\boldsymbol{t}}$. Hence it is sufficient to show that $S_j - S$ is uniformly $o(n^{1/2+\varepsilon})$ with probability greater than $e^{-n^{6\varepsilon/5}}$ in $\mathcal{G}_{\boldsymbol{t}}$. We shall use Theorem 3.1, a concentration result due to Chernoff.

Fix a *white* vertex u_j. For $1 \leq k \leq n$ we consider the random variable $X_{j,k}$ such that $X_{j,k} = 1$ if there is an edge between u_j and v_k and zero otherwise. In our model $\mathcal{G}_{\boldsymbol{t}}$ the degrees of the v_k's are prescribed and any bipartite graph satisfying those degrees is equally likely. Hence the t_k edges emanating from vertex v_k have an equal chance of joining to any set of t_k different vertices from u_1 to u_m. Therefore,

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(X_{j,k} = 1) = t_k/m. \tag{6.1}$$

By the definition of our random variable, $S_j = \sum_k X_{j,k}$. Note also that the expected value of S_j is λn by the following calculation,

$$E(S_j) = \textstyle\sum_k P(X_{j,k} = 1) = m^{-1} \sum_k T_k = m^{-1}\lambda mn = \lambda n.$$

The $X_{j,k}$ are independent random variables, because the edges incident with one *black* vertex do not affect the probability that edges will be incident with any other *black* vertex.

We also note that since $0 \leq S_j \leq n$ we have $0 \leq E(S_j) \leq n$. (This is a very crude bound but it suffices for this lemma.) Hence by Theorem 3.1, (letting $k = n^{1/2+3\varepsilon/4}$),

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(|S_j - \lambda n| \geq n^{1/2+3\varepsilon/2}) < 2e^{\frac{-n^{1+3\varepsilon/2}}{2(E(X)+n^{1/2+3\varepsilon/2}/3)}} < e^{-n^{5\varepsilon/4}}.$$

Hence,

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(\forall j, \quad |S_j - \lambda n| < n^{1/2+3\varepsilon/2}) \geq 1 - me^{-n^{5\varepsilon/4}} \geq 1 - e^{-n^{6\varepsilon/5}}, \tag{6.2}$$

and we have our result. □

6.2 *Pathological* degree sequences in $\mathcal{G}_{\boldsymbol{t}}$

6.2.1 *Locally ordered* bipartite graphs

We will define a new probability space, $\mathcal{G}_{\boldsymbol{t}}^{\varnothing}$, over specially labelled bipartite graphs on (m, n) vertices with *black* degree sequence $\boldsymbol{t}$. We will show that the probability that a random graph has any particular degree sequence $(\boldsymbol{s}, \boldsymbol{t})$ in $\mathcal{G}_{\boldsymbol{t}}$ and $\mathcal{G}_{\boldsymbol{t}}^{\boldsymbol{a}}$. Hence bound the probability of non-*pathological* degree sequences in $\mathcal{G}_{\boldsymbol{t}}$ we instead prove the corresponding bounds in $\mathcal{G}_{\boldsymbol{t}}^{\varnothing}$.

In order to define the probability space, $\mathcal{G}_{\boldsymbol{t}}^{\varnothing}$, we define an allowable labelling for the graphs in the domain of $\mathcal{G}_{\boldsymbol{t}}^{\varnothing}$. We term these allowable labellings *locally ordered*. The name reflects the relation to *edge ordered* labellings which are well studied in the literature, see, for example [GK73].

Definition 6.1 (edge-ordering)**.** *An* ***edge ordering*** *L of a graph G is a bijection from the edges of G to $\{1, 2, \ldots, |E(G)|\}$.*

Definition 6.2 (*locally ordered*)**.** *A bipartite graph G on (m, n) vertices together with an edge-labelling L is said to be* **locally ordered** *if, for each $1 \leq k \leq m$, the subgraph (and sub-labelling) $G \backslash \{v_l : l \neq k\}$ is* edge ordered*.*

We will often write the *locally ordered* graph $G^o = (G, L)$ to refer to the graph G together with the *locally ordered*labelling L.

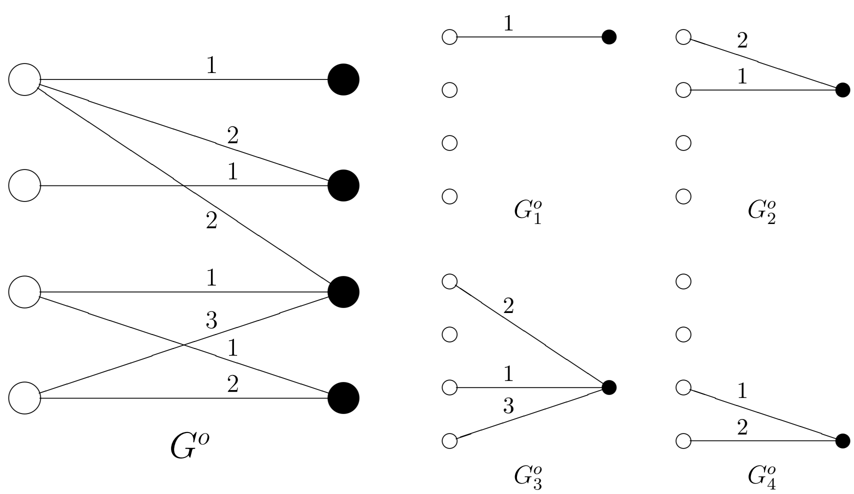


Figure 6.1: Here, G^o is *locally ordered* because G^o_1, G^o_2, G^o_3 and G^o_4 are *edge-ordered*

6.2.2 Decision tree

For each triple of parameters $\boldsymbol{t}, m, n$, we will construct a rooted tree. In this tree, the root node corresponds to the set of all graphs on (m, n) vertices with *black* degree sequence $\boldsymbol{t}$. Also, the leaves of the tree each correspond to single graphs on (m, n) vertices with *black* degree sequence $\boldsymbol{t}$; one leaf for each such graph.

We illustrate the decision tree with *black* degree sequence $\boldsymbol{t} = (1, 1, 1)$ and vertices $(m, n) = (3, 3)$ in Figure 6.4. To define the structure of a decision tree in general requires the following quagmire of definitions.

We define sets of tuples. Below, S_{t_k} is the group of permutations of the numbers from 1 to t_k. We also write $M = \sum_k t_k$.

Definition 6.3 $(\mathcal{A}, \mathcal{A}_l)$**.**

$$\mathcal{A} := \mathcal{A}(\boldsymbol{t}) = \left\{ \boldsymbol{a} \ : \ |\boldsymbol{a}| = M \ \ \& \ \ \forall k, \ \ (a_{\sum_{h<k} t_h + 1}, \ldots, a_{\sum_{h\le k} t_h}) \in S_{t_k} \right\}$$
$$\mathcal{A}_l := \mathcal{A}_l(\boldsymbol{t}) = \left\{ \ (a_1, \ldots, a_l) \mid (a_1, \ldots, a_M) \in \mathcal{A}(\boldsymbol{t}) \right\}$$

Definition 6.4 (reference function $\mathcal{R}$)**.** *This is defined for a n-tuple,* $\boldsymbol{t}$*, and integer* $1 \le l \le n$*. We define,*

$$\mathcal{R}(l, \boldsymbol{t}) := (h, i) \ \textit{where} \ l = \sum_{k<h} t_k + i.$$

Definition 6.5 $(Y_l^{\boldsymbol{a}})$**.** *Let* l*-tuple* $\boldsymbol{a} \in \mathcal{A}_l$ *and for each* $1 \le i \le l$*, let* $(h_i, r_i) = \mathcal{R}(i, \boldsymbol{t})$*. Then define* $Y_l^{\boldsymbol{a}} = Y_l^{\boldsymbol{a}}(m, n, \boldsymbol{t})$ *by,*

$$Y_l^{\boldsymbol{a}}(m, n, \boldsymbol{t}) := \left\{ G \in \mathcal{B}^o_{m,n,\boldsymbol{t}} : \forall \ 1 \le i \le l, \ \ (v_{h_i}, u_{r_i}, a_i) \in E^o(G) \right\}.$$

Each set of graphs $Y_l^{\boldsymbol{a}}$, corresponds to a node at level l in our decision tree. In Figure 6.2 we give some examples that relate to the decision tree in Figure 6.4 on p60.

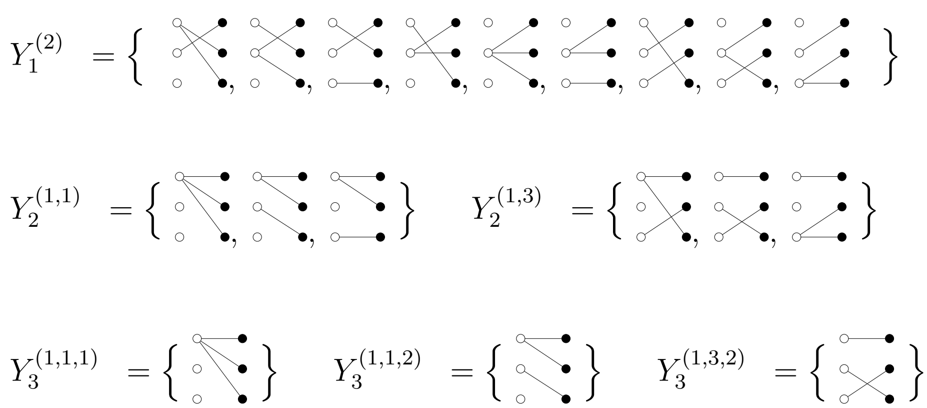


Figure 6.2: We illustrate the sets $Y_l^{\boldsymbol{a}} = Y_l^{\boldsymbol{a}}(3, 3, (1, 1, 1))$ for various l-tuples $\boldsymbol{a}$. Each $Y_l^{\boldsymbol{a}}$ is a subset of the set of all *locally ordered* bipartite graphs on $(3, 3)$ vertices with *black* degree sequence $\boldsymbol{t}$. (All edges shown are labelled '1'.)

Decision tree construction for $m, n, \boldsymbol{t}$**.**

- *nodes*

  For each $0 \le l \le M$, the set of nodes at level l in the tree is $\{Y_l^{\boldsymbol{a}}\}_{\boldsymbol{a} \in \mathcal{A}_l}$.

- *edges*

  Let $\boldsymbol{a} \in \mathcal{A}_l$ for some $1 \leq l \leq M$. Set $\boldsymbol{a}' = (a_1, \ldots, a_{l-1})$ and define the *parent* of $\{Y_l^{\boldsymbol{a}}\}$ to be $\{Y_{l-1}^{\boldsymbol{a}'}\}$. An edge is drawn between each node and its parent node.

Note that if the l-tuple $\boldsymbol{a}$ and $(l+1)$-tuple $\boldsymbol{b}$ agree in their first l elements, then $Y_{l+1}^{\boldsymbol{b}} \subset Y_l^{\boldsymbol{a}}$. Furthermore for the l-tuple $\boldsymbol{a}$ define $C_l^{\boldsymbol{a}} := \{\boldsymbol{b} \in \mathcal{A}_{l+1} \ : \ (b_1, \ldots, b_l) = \boldsymbol{a}\}$. Then the sets $\{Y_{l+1}^{\boldsymbol{b}}\}_{\boldsymbol{b} \in C_l^{\boldsymbol{a}}}$ partition $Y_l^{\boldsymbol{a}}$. Hence the nodes defined form a decision tree as in [CL06, p.106].

We adopt the usual language of decision trees. For each $\boldsymbol{b} \in C_l^{\boldsymbol{a}}$ we call the node $Y_{l+1}^{\boldsymbol{b}}$ a *child* of the node $Y_l^{\boldsymbol{a}}$. Also, two $(l+1)$-tuples $\boldsymbol{b}, \boldsymbol{b}' \in C_l^{\boldsymbol{a}}$ are termed *sibling* nodes.

6.2.3 Graph *ordered-half*-model, $\mathcal{G}_{\boldsymbol{t}}^{\boldsymbol{a}}$

Definition 6.6 ($\mathcal{B}^o_{m,n,\boldsymbol{t}}$)**.**

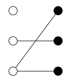
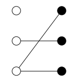
$$\mathcal{B}^o_{m,n,\boldsymbol{t}} := \{\text{locally ordered } \textit{bipartite graphs on } (m,n) \textit{ vertices with } \text{black } \textit{degree sequence } \boldsymbol{t}\}.$$

Definition 6.7 (Graph *ordered-half*-model, $\mathcal{G}_{\boldsymbol{t}}^{\boldsymbol{a}}$)**.** The **graph *ordered-half*-model** $\mathcal{G}_{\boldsymbol{t}}^{\boldsymbol{a}} = \mathcal{G}(m, n, \boldsymbol{a}, \boldsymbol{t})$ *has domain* $\mathcal{B}^o_{m,n,\boldsymbol{t}}$*. Its support is the subset of these graphs,* $Y_{|a|}^{\boldsymbol{a}}(m, n, \boldsymbol{t})$*. All graphs in the support have equal probability.*

When $\boldsymbol{a} = \varnothing$ the only restriction on the edges in the support of $\mathcal{G}_{\boldsymbol{t}}^{\boldsymbol{a}}$ is that the *black* degree sequence is $\boldsymbol{t}$. This is very similar to the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$, except in $\mathcal{G}_{\boldsymbol{t}}^{\varnothing}$ all edges have labels. Every graph in the support of $\mathcal{G}_{\boldsymbol{t}}$ corresponds to the same number or labelled graphs in $\mathcal{G}_{\boldsymbol{t}}^{\varnothing}$. Hence, for any bipartite graph G with *black* degree sequence $\boldsymbol{t}$,

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(G) = \mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}(\{G^o = (G, L) : \text{ for some } \textit{locally ordered} \text{ labelling } L\}). \tag{6.3}$$

We illustrate some example calculations in $\mathcal{G}_{\boldsymbol{t}}^{\boldsymbol{a}}$ in Figure 6.3.

$\mathbb{P}_{\mathcal{G}^{(3)}_{(1,1,1)}}\Big($  $\Big) = \frac{1}{9}$ $\quad$ $\mathbb{P}_{\mathcal{G}^{(2)}_{(1,1,1)}}\Big($  $\Big) = 0$

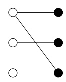
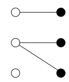
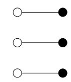
$\mathbb{P}_{\mathcal{G}^{(1,2)}_{(1,1,1)}}\Big($  $\Big) = \mathbb{P}_{\mathcal{G}^{(1,2)}_{(1,1,1)}}\Big($  $\Big) = \mathbb{P}_{\mathcal{G}^{(1,2)}_{(1,1,1)}}\Big($  $\Big) = \frac{1}{3}$

Figure 6.3: We work in the probability spaces $\mathcal{G}^{\boldsymbol{a}}_{\boldsymbol{t}}$ for $\boldsymbol{t} = (1,1,1)$ and various tuples $\boldsymbol{a}$. The domain in each case is the set of all *locally ordered* bipartite graphs on $(3,3)$ vertices with *black* degree sequence $(1,1,1)$. In the diagrams each edge has label '1'.

Definition 6.8 (near-regular)**.** *An* locally ordered *bipartite graph, G, is called* near-regular*, (abbreviated* nreg*) if it satisfies,*

$$\forall j, \quad |s_j(G) - \lambda n| < n^{1/2+3\varepsilon/2}.$$

Corollary 6.9. *Fix $a \in \mathbb{R}^+$ and $\varepsilon > 0$ such that $a < \frac{1}{2}$. Let $\mathcal{G}^{\varnothing}_{\boldsymbol{t}} = \mathcal{G}^{\varnothing}_{\boldsymbol{t}}(m,n)$ be as in Definition 6.7. Then as $m, n \to \infty$ subject to Conditions 6.1,*

$$\mathbb{P}_{\mathcal{G}^{\varnothing}_{\boldsymbol{t}}}(G \text{ is nreg}) \geq 1 - e^{-n^{6\varepsilon/5}}.$$

Proof. We have earlier observed, see (6.3), that any event on the degrees of a graph has the same probability in $\mathcal{G}_{\boldsymbol{t}}$ and $\mathcal{G}^{\varnothing}_{\boldsymbol{t}}$. Now, note that by Definition 6.8 of a *near-regular* graph G, that the statement we are required to prove is precisely (6.2). So we are done. $\square$

6.2.4 Edge-uncovering martingale

Fix $0 \leq l \leq M$, then the sets $\{Y^{\boldsymbol{a}}_l\}_{\boldsymbol{a} \in \mathcal{A}_l}$ partition our domain, $\mathcal{B}_{m,n,\boldsymbol{t}}$. Hence, the power set of these partitions,

$$\mathcal{F}_l := \{Y^{\boldsymbol{a}}_l\}_{\boldsymbol{a} \in \mathcal{A}_l} \tag{6.4}$$

is a σ-algebra.

We are now in a position to define our martingale. For each $0 \leq i \leq M$, we define the random variable,

$$X_i = \mathbb{E}_{\mathcal{G}^{\varnothing}_{\boldsymbol{t}}}\big(\sum_j (S_j - \lambda n)^2 \mid \mathcal{F}_i\big). \tag{6.5}$$

In the next lemma we show that these random variables do indeed form a martingale.

Lemma 6.10. *The random variables $X_0, \ldots, X_M$ as defined in* (6.5) *form a martingale.*

Proof. Notice that $\left(\sum_j (S_j - \lambda n)^2\right)^2 < (mn^2)^2 < \infty$. This implies that each $E(X_l)^2 < \infty$ for each $0 \leq l \leq M$. Hence by the Definition 3.3 of martingales it remains now to show,

$$\mathbb{E}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(X_{l+1} \mid \mathcal{F}_l\big) = X_l. \tag{6.6}$$

We show (6.6) by the following,

$$\begin{aligned}
\mathbb{E}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(X_{l+1} \mid \mathcal{F}_l\big) &= \mathbb{E}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(\mathbb{E}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}(\sum_j (S_j - \lambda n)^2 \mid \mathcal{F}_{l+1}) \mid \mathcal{F}_l\big) \\
&= \mathbb{E}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}(\sum_j (S_j - \lambda n)^2 \mid \mathcal{F}_l\big) \qquad (6.7) \\
&= X_l.
\end{aligned}$$

The first and last lines follow by the definitions of X_{l+1} and X_l respectively. Because $\mathcal{F}_l \subset F_{l+1}$, line (6.7) follows by the tower of expectation property (see Lemma 3.5). $\square$

In the next section we will find a bound for $|X_i - X_{i+1}|$. Actually, we find two bounds, a weaker one that holds at all points in the domain and stronger one that holds for most of the domain. This will then allow us to use the generalised Azuma Theorem, (our Theorem 3.8), to bound the difference, $|X_0 - X_M|$.

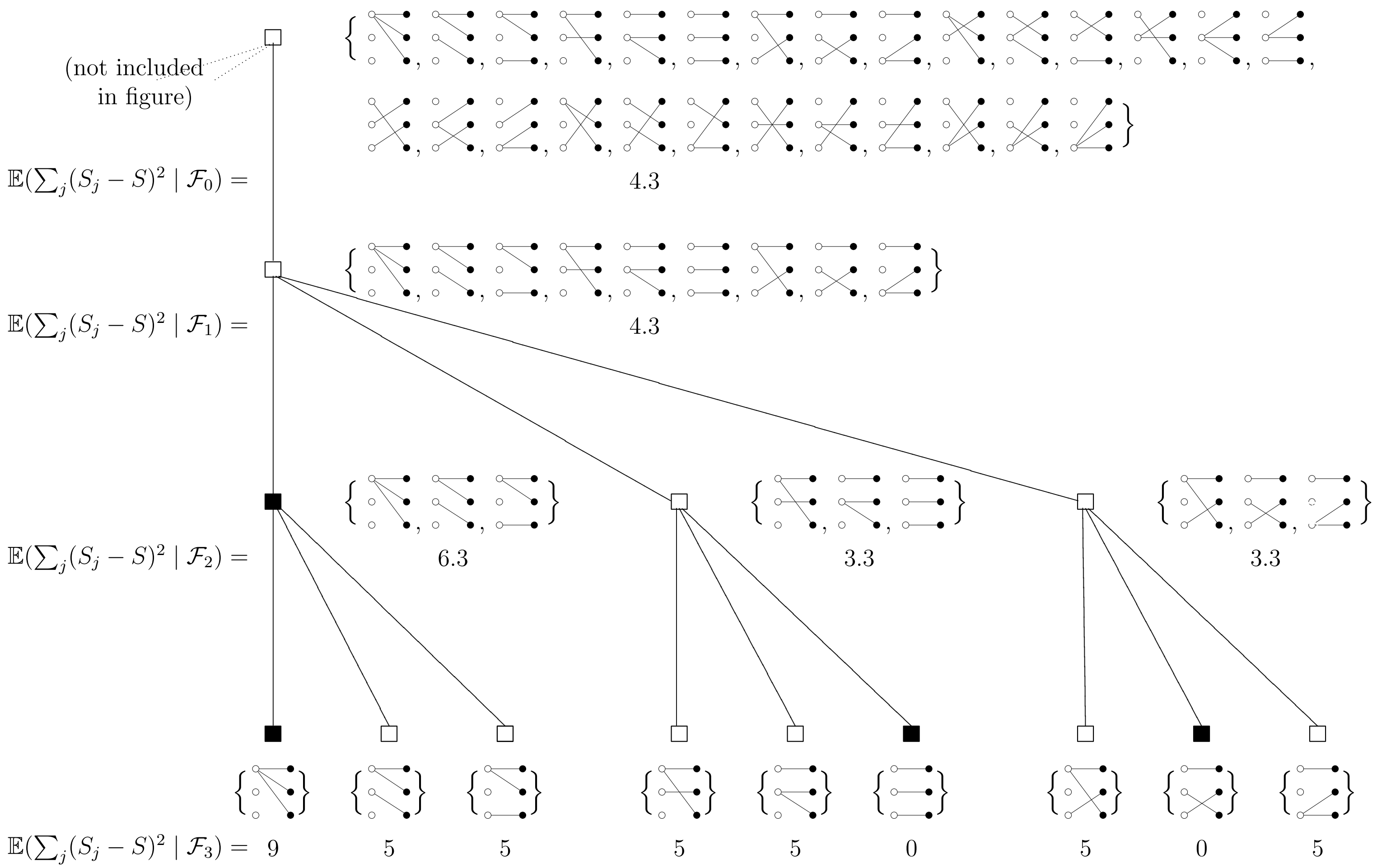


Figure 6.4: The decision tree for bipartite graphs on $(3, 3)$ vertices with *black* degree sequence $(1, 1, 1)$.

Let $X_i = \mathbb{E}(\sum_j (S_j - S)^2 | \mathcal{F}_i)$. For each node at level i in the tree we display the value of $X_i(G)$ for graphs G at that node. We have shaded the nodes at level i for which $|X_{i+1}(G) - X_i(G)| \geq 2$ for graphs G at that node (such nodes are referred to to as *bad* in the text). Note this is not the same cut-off we use in our calculations.

6.2.5 Bounding Arguments

Definition 6.11 (toxic)**.** *Fix $0 \leq l \leq M$. Then l-tuple $\boldsymbol{a} \in \mathcal{A}_l$ is* toxic *if*

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(G \text{ not near-regular} \mid G \in Y_l^{\boldsymbol{a}}\big) > n^{-1/2}.$$

Lemma 6.12. *Fix $0 \leq l \leq M$, $0 < a < \frac{1}{2}$ and $\varepsilon > 0$. Let $\mathcal{G}_{\boldsymbol{t}} = \mathcal{G}_{\boldsymbol{t}}(m,n)$ be as in Definition 6.7. Then as $m, n \to \infty$ subject to Conditions 6.1,*

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(\{Y_l^{\boldsymbol{a}} : \boldsymbol{a} \text{ toxic}\}\big) < n^{1/2} e^{-n^{3\varepsilon/2}}.$$

Proof. This proof proceeds by contradiction. Assume that

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(\{Y_l^{\boldsymbol{a}} : \boldsymbol{a} \text{ toxic}\}\big) \geq n^{1/2} e^{-n^{3\varepsilon/2}}.$$

Then,

$$\begin{aligned}
\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(G \text{ not nreg}\big) &= \sum_{\boldsymbol{a} \in \mathcal{A}_l} \mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(G \text{ not nreg } \mid G \in Y_l^{\boldsymbol{a}}\big) \mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(Y_l^{\boldsymbol{a}}\big) && (6.8)\\
&\geq \sum_{\text{toxic } \boldsymbol{a} \in \mathcal{A}_l} \mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(G \text{ not nreg} \mid G \in Y_l^{\boldsymbol{a}}\big) \mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(Y_l^{\boldsymbol{a}}\big) && (6.9)\\
&\geq n^{-1/2} \sum_{\text{toxic } \boldsymbol{a} \in \mathcal{A}_l} \mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(Y_l^{\boldsymbol{a}}\big) && (6.10)\\
&> e^{-n^{3\varepsilon/2}} && (6.11)
\end{aligned}$$

We justify the calculation above line by line. Line (6.8) follows by Bayes rule and because the sets $\{Y_l^{\boldsymbol{a}}\}_{\boldsymbol{a} \in \mathcal{A}_l}$ partition all graphs, $\mathcal{B}_{m,n,\boldsymbol{t}}$, in the domain of $\mathcal{G}_{\boldsymbol{t}}^{\varnothing}$. Line (6.9) then follows because all probabilities are non-negative. Definition 6.11 then implies (6.10). The last line, (6.11) follows by our assumptive hypothesis.

To complete the proof, observe that (6.11) contradicts Corollary (6.9). □

Definition 6.13 (bad)**.** *Fix $1 \leq l \leq M$. Then l-tuple $\boldsymbol{a} \in \mathcal{A}_l$ is* bad *if $Y_l^{\boldsymbol{a}}$ is toxic or has a sibling node which is toxic.*

Lemma 6.14. *Fix $0 \leq l \leq M$, $0 < a < \frac{1}{2}$ and $\varepsilon > 0$. Let $\mathcal{G}_{\boldsymbol{t}} = \mathcal{G}_{\boldsymbol{t}}(m,n)$ be as in Definition 6.7. Then as $m, n \to \infty$ subject to Conditions 6.1,*

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(\{Y_l^{\boldsymbol{a}} : \boldsymbol{a} \text{ bad}\}\big) < n^{3/2+\varepsilon} e^{-n^{3\varepsilon/2}}.$$

Proof. This follows because there at most m children to any node. Hence,

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(\{Y_l^{\boldsymbol{a}} : \boldsymbol{a} \text{ bad}\}\big) < m\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(\{Y_l^{\boldsymbol{a}} : \boldsymbol{a} \text{ toxic}\}\big) < n^{3/2+\varepsilon} e^{-n^{3\varepsilon/2}}.$$

□

Bijection between sibling nodes, $\phi^{G \to H}$ Consider the two sets of bipartite graphs: $Y_l^{\boldsymbol{a}}$ and $Y_l^{\boldsymbol{b}}$, where $|a| = |b| = l$. When $\boldsymbol{a}$ and $\boldsymbol{b}$ differ only in the last element we expect the graphs in the two sets to be somewhat similar.[1]

We make this idea precise by defining a bijection ϕ from each graph in $Y_l^{\boldsymbol{a}}$ to one in $Y_l^{\boldsymbol{b}}$ with symmetric difference at most four labelled edges. The intuitive idea of the bijection ϕ is we want to match each $H \in Y_l^{\boldsymbol{a}}$ with the closest resembling $H' \in Y_l^{\boldsymbol{b}}$.

Definition 6.15 (graph match bijection $\phi_l^{\boldsymbol{a} \to \boldsymbol{b}}$)**.** $\phi_l^{\boldsymbol{a} \to \boldsymbol{b}} = \phi_l^{\boldsymbol{a} \to \boldsymbol{b}}(\boldsymbol{t}, m, n) : Y_l^{\boldsymbol{a}} \to Y_l^{\boldsymbol{b}}$.

The map is defined when $Y_l^{\boldsymbol{a}}$, $Y_l^{\boldsymbol{b}}$ *are both subsets of* locally ordered *bipartite graphs on* (m, n) *vertices with* black *degree sequence* $\boldsymbol{t}$ *and* $a_h = b_h$ *for all* $1 \leq h < l$.

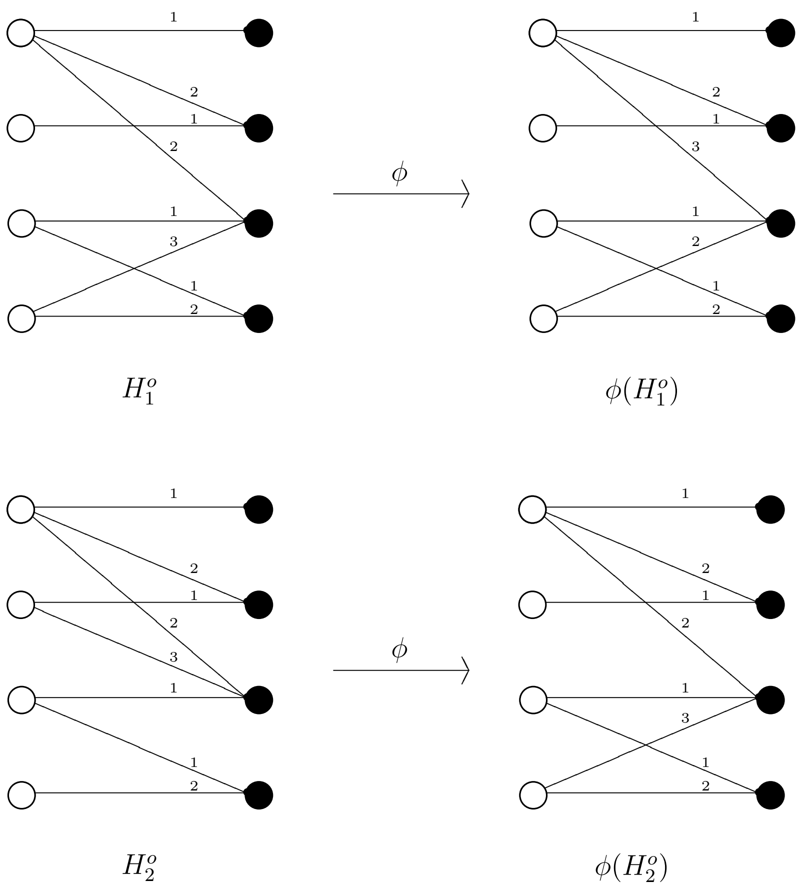
Let $(h, i) = f_{\boldsymbol{t}}(l)$ *and fix* $G \in Y_l^{\boldsymbol{a}}$. *We specify the graph* $\phi(G)$ *by giving the symmetric difference* $\triangle$ *between the labelled edge sets of* G *and* $\phi(G)$. *There are two possibilities for edge labellings of* G *which give the following two symmetric differences:*

if $(v_h, v_{b_i}, r) \in E(G^o)$ *(for some* $r > i$*),*
then $E^o(G^o) \triangle E^o(\phi(G^o)) = \{(u_{a_i}, v_k, i), (u_{b_i}, v_k, r), (u_{a_i}, v_k, r), (u_{b_i}, v_k, i)\}$

else,
$E^o(G^o) \triangle E^o(\phi(G^o)) = \{(u_{a_i}, v_k, i), (u_{b_i}, v_k, i)\}$.

We shall give an example. Consider *locally ordered* bipartite graphs on (4,4) vertices, with *black* degree sequence $\boldsymbol{t}$ =(1,2,3,2). We illustrate $\phi = \phi(4, 4, (1, 2, 3, 2))$ by displaying the two graphs H_1^0 and H_2^0 before and after the map ϕ is applied. See Figure 6.5.

[1] In the language of decision trees, this is equivalent to saying $Y_l^{\boldsymbol{a}}$ and $Y_l^{\boldsymbol{b}}$ are sibling nodes. We define a bijection sending graphs at a node into any sibling node.



$$E^o(H_1^o) \triangle E^o(\phi(H_1^o)) = (u_1, v_3, 2) + (u_4, v_3, 3) + (u_1, v_3, 3) + (u_4, v_3, 2)$$

$$E^o(H_2^o) \triangle E^o(\phi(H_2^o)) = (u_2, v_3, 3) + (u_4, v_3, 3)$$

Figure 6.5: In the illustration the bijection map ϕ is, $\phi = \phi_5^{((1),(2,1),(3,1))\to((1),(2,1),(3,4))} : Y_5^{((1)(2,1)(3,1))} \to Y_5^{((1)(2,1)(3,4))}$.

Bounds between a node and its child

Lemma 6.16. *Let G be an* locally ordered *bipartite graph on (m, n) vertices with* black *degree sequence $\boldsymbol{t}$. Fix a* white *vertex u_j.*

1. *Then,*
$$|\sum_j (s_j(G) - \lambda n)^2 - \sum_j (s_j(\phi(G)) - \lambda n)^2| < 2n + 2.$$

2. *Additionally, suppose that G is near-regular, then,*
$$|\sum_j (s_j(G) - \lambda n)^2 - \sum_j (s_j(\phi(G)) - \lambda n)^2| < 2n^{1/2+\varepsilon} + 2.$$

Proof. By property of ϕ either both G and $\phi(G)$ have the same edge set (with two pairs of labels interchanged) or G can be obtained from $\phi(G)$ by moving one edge. In the first case our lemma is trivially true, so we assume the second. Hence for some $1 \leq r, q \leq m$ and some $1 \leq k \leq n$, we have $E(G) \triangle E(\phi(G)) = (u_r, v_k) + (u_q, v_k)$. Assume, w.l.o.g that $(u_r, v_k) \in E(G)$, then,

$$\begin{aligned}
&\sum_j (s_j(G) - \lambda n)^2 - \sum_j (s_j(\phi(G)) - \lambda n)^2 \\
&\quad = (s_r(G) - \lambda n)^2 + (s_q(G) - \lambda n)^2 - (s_r(\phi(G)) - \lambda n)^2 - (s_q(\phi(G)) - \lambda n)^2 \\
&\quad = (s_r(G) - \lambda n)^2 + (s_q(G) - \lambda n)^2 - (s_r(G) - 1 - \lambda n)^2 + (s_q(G) + 1 - \lambda n)^2 \\
&\quad = 2(s_r(G) - \lambda n) - 2(s_q(G) - \lambda n) + 2 \qquad (6.12)
\end{aligned}$$

Note that (6.12) can be simplified further. Taking absolute values,

$$|2(s_r(G) - \lambda n) - 2(s_q(G) - \lambda n) + 2| = |2(s_r(G) - s_q(G)) + 2| \leq 2n + 2. \qquad (6.13)$$

The bound (6.13) follows because the degrees of any two *white* vertices must lie between zero and n and hence their absolute difference is less than or equal to n. This proves part a.

When we make the additional assumptions in (2) we can see that the last line (6.12) must be less than $2n^{1/2+\varepsilon} + 2$. This follows because we required, in G, that $|(S_j(G) - \lambda n)| < n^{1/2+\varepsilon}$ for each j. This completes the proof of (2).

$\square$

For the following proof it will be convenient to define $P^{i+1}(G) = \mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}(H) / \sum_{J \equiv_i H} \mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}(J)$. This notation is consistent with that used in [SS87], on which we have modelled the method of our proof.

Lemma 6.17.

Let $X : \Omega \to \mathbb{R}$ be the random variable, $X(G) := \sum_j (S_j(G) - \lambda n)^2$. For $0 \leq i \leq M$ let X_i be the random variable defined in (6.5).

1. *Then,*
$$|X_{i+1}(G) - X_i(G)| < 2n + 2.$$

2. *Additionally, suppose that we have a **good** i-tuple $\boldsymbol{a}$. Then for $G \in Y_i^{\boldsymbol{a}}$,*

$$|X_{i+1}(G) - X_i(G)| < n^{1/2+2\varepsilon}$$

Proof. (of 2.) We prove the more restricted case first.

Fix $G \in Y_i^{\boldsymbol{a}}$. Then *doing average of averages*,[2]

$$X_i(G) = \sum_{H \equiv_i G} X_{i+1}(H) P^i(H) \tag{6.14}$$

$$\therefore X_{i+1}(G) - X_i(G) = \sum_{H \equiv_i G} \big(X_{i+1}(G) - X_{i+1}(H)\big) P^i(H). \tag{6.15}$$

Fix $H \equiv_i G$, then, similarly to (6.14):

$$X_{i+1}(G) = \sum_{J \equiv_{i+1} G} X(J) P^{i+1}(J) \tag{6.16}$$

$$X_{i+1}(H) = \sum_{K \equiv_{i+1} H} X(K) P^{i+1}(K). \tag{6.17}$$

We want to bound the difference between (6.16) and (6.17). Consider the bijection:

$$\phi = \phi_{i+1}^{G \to H} : \{J \ : \ J \equiv_{i+1} G\} \to \{K \ : \ K \equiv_{i+1} H\}.$$

Also note by the definition of the nodes in the decision tree, $P^{i+1}(J) = P^{i+1}(\phi(J))$. Hence can apply this bijection to obtain

$$X_{i+1}(G) - X_{i+1}(H) = \sum_{J \equiv_{i+1} G} \big(X_{i+1}(J) - X_{i+1}(\phi(J))\big) P^{i+1}(J). \tag{6.18}$$

By Lemma 6.16, for each *near-regular* $J \equiv_{i+1} G$ we have,

$$\big|X_{i+1}(J) - X_{i+1}(\phi(J))\big| < 2n^{1/2+\varepsilon} + 2, \tag{6.19}$$

[2]Quote from Shamir and Spencer, this theorem is somewhat analogous to Theorem 5 in [SS87] from which the quote originates. We follow their method quite closely. Note that Shamir and Spencer find a single bound for all G whereas we find bounds for two separate cases. In their case the random variable $X(G)$ is defined to be the chromatic number of G, for more information on the proof by Shamir and Spencer refer to Section 3.2.4 in this thesis.

and, for each *not near-regular* $J \equiv_{i+1} G$ we have,

$$\big|X_{i+1}(J) - X_{i+1}(\phi(J))\big| < 2n + 2. \tag{6.20}$$

By (6.18):

$$\begin{aligned}
&\big|X_{i+1}(G) - X_{i+1}(H)\big| \\
&\qquad = \sum_{\text{nreg } J\equiv_{i+1}G} \big(X_{i+1}(J) - X_{i+1}(\phi(J))\big)P^{i+1}(J) \\
&\qquad + \sum_{\text{not nreg } J\equiv_{i+1}G} \big(X_{i+1}(J) - X_{i+1}(\phi(J))\big)P^{i+1}(J) \\
&\qquad = \big(2n^{1/2+\varepsilon} + 2\big) \sum_{\text{nreg } J\equiv_{i+1}G} P^{i+1}(J) \\
&\qquad + \big(2n+2\big) \sum_{\text{not nreg } J\equiv_{i+1}G} P^{i+1}(J) &(6.21)\\
&\qquad = \big(2n^{1/2+\varepsilon} + 2\big)\big(1 - n^{-1/2}\big) + \big(2n+2\big)n^{-1/2} &(6.22)\\
&\qquad < n^{1/2+2\varepsilon} &(6.23)
\end{aligned}$$

We justify the above line by line. The first line follows directly from (6.18) and the definition of *near-regular* graphs. Noting the bounds in (6.19) and (6.20) for *near-regular* and *general* graphs respectively yields (6.21). Then (6.22) follows by Lemma 6.14.

So then by (6.15),

$$\big|X_i(G) - X_{i+1}(G)\big| < n^{1/2+2\varepsilon}$$

which yields the desired result. □

Proof. (of 1.) The proof of this case follows the above proof exactly excepting that do not assume any $J \equiv_{i+1} G$ are *near-regular*. So we have a weaker bound. For a general graph G from Lemma 6.16(1) we have, for each $J \equiv_{i+1} G$,

$$\big|X_{i+1}(J) - X_{i+1}(\phi(J))\big| < 2n + 2.$$

And hence we have

$$\big|X_i(G) - X_{i+1}(G)\big| < 2n + 2,$$

as required. □

Lemma 6.18. *Fix* $0 \le l \le M$, $0 < a < \frac{1}{2}$ *and* $\varepsilon > 0$. *Let* $\mathcal{G}_{\boldsymbol{t}} = \mathcal{G}_{\boldsymbol{t}}(m,n)$ *be as in Definition 6.7. For* $1 \le j \le m$, *let* S_j *be the random variable that returns the* degree *of the* white *vertex* u_j.

Then as $m, n \to \infty$ *subject to Conditions 6.1,*

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}\Big(\Big|\frac{\sum_j (S_j - \lambda n)^2 - \mathbb{E}_{\mathcal{G}_{\boldsymbol{t}}}\big(\sum_j (S_j - \lambda n)^2\big)}{\lambda(1-\lambda)mn}\Big| > n^{-1/2+7\varepsilon}\Big) < e^{-n^{3\varepsilon/2}}.$$

Proof. Observe first that by Lemma 4.5, $\lambda(1-\lambda) > \frac{1}{\ln n} > n^{-\varepsilon/2}$, and so $\frac{1}{\lambda(1-\lambda)mn} > n^{-2+2\varepsilon}$. Thus it is sufficient to prove,

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}\Big(\Big|\sum_j (S_j - \lambda n)^2 - \mathbb{E}_{\mathcal{G}_{\boldsymbol{t}}}\big(\sum_j (S_j - \lambda n)^2\big)\Big| > n^{3/2+5\varepsilon}\Big) < e^{-n^{3\varepsilon/2}}. \tag{6.24}$$

Let $c = n^{1/2+2\varepsilon}$ and $\eta = n^{7\varepsilon/2}e^{-6\varepsilon/5}$. To prove (6.24) we will show that the random variables, $X_0, \ldots, X_M$ form a near-c-Lipschitz martingale with exceptional probability η.

By Lemma 6.10, we already have that $X_0, \ldots, X_M$ form a martingale.

Let $G \in Y_l^{\boldsymbol{a}}$ for some **good** node $Y_l^{\boldsymbol{a}}$. Then $|X_i - X_{i+1}|(G) < c$. This is shown in Lemma 6.17. Hence for a fixed l we can bound the set of graphs $\{G'\}$ for which $|X_i - X_{i+1}|(G') \ge c$, by taking the union of all *bad* nodes at level l of the tree. So we bound the sum of probability of all bad nodes over all levels, l, of the decision tree,

$$\sum_{l=1}^{M} \mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(\{Y_l^{\boldsymbol{a}} \,:\, \boldsymbol{a} \text{ bad}\}\big) < Mn^{3\varepsilon/2}e^{-n^{6\varepsilon/5}} < n^{7\varepsilon/2}e^{-6\varepsilon/5} \tag{6.25}$$

In (6.25) the first inequality follows by Lemma 6.14 and the second inequality follows because $M \le mn$ and $m = o(n^{1+\varepsilon})$. This shows our values for η and c hold and so $X_0, \ldots, X_n$ form a near-c-Lipschitz martingale with exceptional probability η.

We can now substitute these values of c and η into the generalised Azuma inequality (Theorem 3.8), to yield (6.24). $\square$

6.2.6 Expectation of $\sum_j (S_j - \lambda n)^2$

The aim of this Section is to show that $(a, b, m, n, \varepsilon)$-*pathological* degree sequences are rare in $\mathcal{G}_{\boldsymbol{t}}$. We already have Lemma 6.1 which shows non-(ε, a)-*regular* sequences are very rare. Hence our current task is showing $\big(1 - \frac{\sum_j (S_j - \lambda n)^2}{\lambda(1-\lambda)mn}\big)\big(1 - \frac{\sum_k (T_k - \lambda m)^2}{\lambda(1-\lambda)mn}\big) = O(n^{-b})$ with high probability. So far we have shown in Lemma 6.18 that with high probability $\frac{\sum_j (S_j - \lambda n)^2 - \mathbb{E}(\sum_j (S_j - \lambda n)^2)}{\lambda(1-\lambda)mn} = O(n^{-b})$. Hence now we want to show that $\frac{\mathbb{E}(\sum_j (S_j - \lambda n)^2)}{\lambda(1-\lambda)mn} =$

$1 + O(n^{-b})$. This is done in Lemma 6.19 below. Before we can prove this, however, we need to calculate some preliminary expectations on our random variables for the *white* vertex degrees: $s_1, \ldots, s_m$.

Fix S_j, we derived earlier in line (6.1) the probability that there is an edge between u_j and v_k is precisely $m^{-1}t_k$. Hence we can write a probability generating function, $B(x)$, for S_j in $\mathcal{G}_{\boldsymbol{t}}$.

$$B(x) = \prod_{k=1}^{n} (\frac{t_k}{m}x + \frac{m - t_k}{m})$$

We can now calculate the expectation of S_j and of S_j^2. By the theory of generating functions explained in Section 3.3 this requires the evaluation of some differential operators on the function $B(x)$. These are slightly messy and so are relegated to the appendix, the following two results are from lines (8.5) and (8.6). Note these hold for all j.

$$\begin{aligned} \mathbb{E}_{\mathcal{G}_{\boldsymbol{t}}}(S_j) &= \left(x\tfrac{d}{dx}B(x)\right)_{x=1} = \lambda n \\ \mathbb{E}_{\mathcal{G}_{\boldsymbol{t}}}(S_j^2) &= \left(x\tfrac{d}{dx}x\tfrac{d}{dx}B(x)\right)_{x=1} = \lambda n + \lambda^2 n^2 - m^{-2}\textstyle\sum_k t_k^2 \end{aligned} \tag{6.26}$$

We are now in a position to prove our result.

Lemma 6.19. *Let $\mathcal{G}_{\boldsymbol{t}} = \mathcal{G}_{\boldsymbol{t}}(m,n)$ be as in Definition 1.8. For $1 \le j \le m$, let S_j be the random variable that returns the* degree *of the* white *vertex u_j. Then as $m, n \to \infty$ subject to Conditions 6.1,*

$$\frac{\mathbb{E}_{\mathcal{G}_{\boldsymbol{t}}}(\sum_j (S_j - \lambda n)^2)}{\lambda(1-\lambda)mn} \quad = \quad 1 + o(n^{-1+5\varepsilon}).$$

Proof. We first calculate the expectation of $\sum_j (S_j - \lambda n)^2$ using the values for expectation from (6.26).

$$\begin{aligned} \mathbb{E}_{\mathcal{G}_{\boldsymbol{t}}}(\sum_j (S_j - \lambda n)^2) &= \mathbb{E}_{\mathcal{G}_{\boldsymbol{t}}}(\sum_j (S_j - \lambda n)^2) \\ &= \mathbb{E}_{\mathcal{G}_{\boldsymbol{t}}}(\sum_j S_j^2) - \lambda^2 mn^2 \\ &= \lambda mn + \lambda^2 mn^2 - m^{-1}\sum_k t_k^2 - \lambda^2 mn^2 \\ &= \lambda mn - m^{-1}\sum_k t_k^2 \end{aligned} \tag{6.27}$$

Fix $0 \le k \le n$, by assumption, $t_k - \lambda m = o(n^{1/2+\varepsilon})$ and so $\sum_k (t_k - \lambda m)^2 = o(n^{2+2\varepsilon})$. By Lemma 4.5, because λ is *acceptable* for a, m and n, $\lambda(1-\lambda) > \ln n$. Hence we have,

$$\frac{\sum_k (t_k - \lambda m)^2}{\lambda(1-\lambda)mn} = o(n^{3\varepsilon}) \tag{6.28}$$

The calculations below use (6.28) to derive a value for $\sum_k t_k^2$,

$$\begin{aligned}
\frac{\sum_k (t_k - m\lambda)^2}{\lambda(1-\lambda)mn} &= o(n^{3\varepsilon}) \\
o(n^{3\varepsilon})\lambda(1-\lambda)mn &= \sum_k (t_k - m\lambda)^2 \\
&= \sum_k (t_k^2) - 2m\lambda \sum_k (t_k) - \lambda^2 m^2 n \\
\Rightarrow \quad \sum_k t_k^2 &= \lambda^2 m^2 n + o(n^{3\varepsilon})\lambda(1-\lambda)mn.
\end{aligned}$$

This approximate value for $\sum_k t_k^2$ can then be used to bound $\frac{\mathbb{E}(\sum_j (S_j - \lambda n)^2)}{\lambda(1-\lambda)mn}$. By inequality (6.27),

$$\begin{aligned}
\frac{\mathbb{E}_{\mathcal{G}_{\boldsymbol{t}}}(\sum_j (S_j - \lambda n)^2)}{\lambda(1-\lambda)mn} &= \frac{\lambda mn - \lambda^2 mn + \lambda(1-\lambda)n}{\lambda(1-\lambda)mn} + o(n^{-3\varepsilon}m^{-1}) \\
&= 1 + m^{-1} + o(n^{-3\varepsilon}m^{-1}) \\
&= 1 + o(n^{-1+5\varepsilon})
\end{aligned}$$

□

This is exactly what we wanted. We can are now ready to show that *pathological* degree sequences are rare in the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$.

Theorem 6.20. *Fix $a, b \in \mathbb{R}^+$ and $\varepsilon > 0$ such that $0 < a + b < \frac{1}{2}$ and $b + 11\varepsilon < 1/2$. Let $\mathcal{G}_{\boldsymbol{t}} = \mathcal{G}_{\boldsymbol{t}}(m, n)$ be as in Definition 1.8.*
For $1 \le j \le m$, let S_j be the random variable that returns the degree *of the* white *vertex u_j.*
Then as $m, n \to \infty$ subject to Conditions 6.1,

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}\big(\, (\boldsymbol{S}, \boldsymbol{T}) \ \text{ is } (a, b, m, n, \varepsilon)\text{-}\textit{pathological} \, \big) \le e^{-n^{7\varepsilon/6}}.$$

Proof. By Lemma 6.1,

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(\ (\boldsymbol{S},\boldsymbol{T})\ \ \textit{is not}\ (\varepsilon,a)\text{-regular}\ ) \le e^{-n^{6\varepsilon/5}}.$$

Hence it is sufficient to show that,

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}\left(\left(1-\frac{\sum_j (S_j-S)^2}{\lambda(1-\lambda)mn}\right)\left(1-\frac{\sum_k (T_k-T)^2}{\lambda(1-\lambda)mn}\right)=O(n^{-b})\right)\ge 1-e^{-n^{3\varepsilon/2}} \tag{6.29}$$

The assumptions of this model imply that the average degree of the *white* vertices is determined by the m-tuple $\boldsymbol{t}$, and so $S=\lambda n$. By two of our previous results, Lemmas 6.19 and 6.18,

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}\left(\left|\frac{\sum_j (S_j-\lambda n)^2}{\lambda(1-\lambda)mn}\right| > n^{-1/2+7\varepsilon}+1+o(n^{-1+5\varepsilon})\right) < e^{-n^{3\varepsilon/2}}. \tag{6.30}$$

Finally, recall by the assumptions of the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$, all random graphs in $\mathcal{G}_{\boldsymbol{t}}$ have *black* degree sequence $\boldsymbol{t}$. Hence, $T=\lambda m$ and for each $1\le k\le n$, $T_k=t_k$. By (6.28),

$$\frac{\sum_k (t_k-\lambda m)^2}{\lambda(1-\lambda)mn}=o(n^{3\varepsilon}). \tag{6.31}$$

Combining (6.30) and (6.31) then yields,

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}\left(\left|\left(1-\frac{\sum_j (S_j-S)^2}{\lambda(1-\lambda)mn}\right)\left(1-\frac{\sum_k (T_k-T)^2}{\lambda(1-\lambda)mn}\right)\right| > n^{-1/2+11\varepsilon}\right)\le e^{-n^{3\varepsilon/2}} \tag{6.32}$$

In particular, because $b+11\varepsilon<1/2$, we get $n^{-1/2+11\varepsilon}=O(n^{-b})$. Thus, (6.32) implies (6.29) and so we are done. $\square$

Part III

New results: Approximation by binomial models

In this section we will define simple binomial probability models in which the probability of an $(m+n)$-tuple $(\boldsymbol{s}, \boldsymbol{t})$ is asymptotically very close to the probability of a degree sequence $(\boldsymbol{s}, \boldsymbol{t})$ occurring in the random graph models.

There are three random bipartite graph models for which we find a corresponding binomial model. These three models are the graph p-model, $\mathcal{G}_p$, the graph *edge*-model, $\mathcal{G}_M$, and the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$. (See Definitions 1.6, 1.7 and 1.8 respectively.)

To achieve this we construct binomial models which we call the binomial p-model, $\mathcal{B}_p$, binomial *edge*-model, $\mathcal{B}_M$, and binomial *half*-model, $\mathcal{B}_{\boldsymbol{t}}$, respectively. The latter two binomial models $\mathcal{B}_M$ and $\mathcal{B}_{\boldsymbol{t}}$ straight away match very closely with $\mathcal{G}_M$ and $\mathcal{G}_{\boldsymbol{t}}$. The first case needs more work and so we define an integrated version of $\mathcal{B}_p$, the binomial *integrated*-model, denoted $\mathcal{V}_p$, which then matches the graph p-model, $\mathcal{G}_p$ very closely.

We construct the binomial probability spaces $\mathcal{B}_p$, $\mathcal{B}_M$ and $\mathcal{B}_{\boldsymbol{t}}$ as restrictions of a common probability space, the binomial *independent*-model, denoted $\mathcal{I}_p$. The *independent*-model is defined in terms of random variables $S_1, \ldots, S_m$ which are independent binomials with parameters (n,p) and random variables $T_1, \ldots, T_n$ which are independent binomials with parameters (m,p). The three binomial models are then defined by considering different subspaces (corresponding to different constraints) of this *independent*-model $\mathcal{I}_p$.

We proceed by first defining our binomial models in Section 7.1. After this we show the correspondences between these binomial models and our random graph models in Section 7.2. The ordering of Section 7.2 reflects the difficulty in obtaining these correspondences and so we defer our discussion the graph p-model, $\mathcal{G}_p$ to Section 7.2.3, after the graph *edge*-model, $\mathcal{G}_M$ and the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$ in Subsections 7.2.1 and 7.2.2 respectively.

Chapter 7

Binomial probability spaces

7.1 Definition of binomial models

7.1.1 Binomial *independent*-model, $\mathcal{I}_p$.

We define our binomial probability models in terms of a common model (with some added constraints). This core model, we call the binomial *independent*-model, $\mathcal{I}_p$.

Definition 7.1 (Binomial *independent*-model, $\mathcal{I}_p$.)**.** The **binomial *independent*-model** $\mathcal{I}_p = \mathcal{I}_p(m,n)$ is an $(m+n)$-dimensional probability space with domain $\{0,1,2,\ldots,n\}^m \times \{0,1,2,\ldots,m\}^n$. For $j = 1,\ldots,m$ and $k = 1,\ldots,n$ we define S_j and T_k to be independent binomially distributed random variables with parameters (n,p) and (m,p) respectively. That is, we have, $\mathbb{P}_{\mathcal{I}_p}(S_j = a) = \binom{n}{a}p^a(1-p)^{n-a}$ and $\mathbb{P}_{\mathcal{I}_p}(T_k = a) = \binom{m}{a}p^a(1-p)^{m-a}$. Our $(m+n)$-dimensional space has the random variables $(\boldsymbol{S} = (S_1,\ldots,S_m), \boldsymbol{T} = (T_1,\ldots,T_n))$.

The **probability of the** $(m+n)$**-tuple,** $(\boldsymbol{s},\boldsymbol{t})$**, in the binomial *independent*-model**, $\mathcal{I}_p$ is given by,

$$\mathbb{P}_{\mathcal{I}_p}(\boldsymbol{S} = \boldsymbol{s}, \boldsymbol{T} = \boldsymbol{t}) = \prod_{j=1}^{m}\binom{n}{s_j}p^{s_j}q^{n-s_j}\prod_{k=1}^{n}\binom{m}{t_k}p^{t_k}q^{m-t_k}.$$

We will proceed to define our other binomial models as probability spaces on a subset of the domain of the *independent*-model, $\mathcal{I}_p$.

7.1.2 Binomial p-model, $\mathcal{B}_p$.

Our constraint for the first binomial model is to require that the only $(m+n)$-tuples $(\boldsymbol{s},\boldsymbol{t})$ with non-zero probability in the space are those for which the sum of the first m

terms is the same as the sum of the last n terms, i.e. $(\boldsymbol{s},\boldsymbol{t}) \in I_{m,n}$. Note this is a sensible constraint. We are hoping to use these binomial models to approximate the random bipartite graph models. The $(m+n)$-tuples in the domain of $\mathcal{B}_p$ will correspond to degree sequences in random bipartite graphs on (m,n) vertices. Recall that the number of edges in a bipartite graph can calculated by summing up the degrees of the vertices on either side. In particular, these two sums, the sums of the degrees of the *white* vertices and the sums of the degrees of the *black* vertices should be equal. This is our motivation for defining the binomial p-model, $\mathcal{B}_p$.

Definition 7.2 (Binomial p-model, $\mathcal{B}_p$.)**.** The **binomial p-model** $\mathcal{B}_p = \mathcal{B}_p(m,n)$ has the same domain as $\mathcal{I}_p$ with support $I_{m,n}$. We formally define the probability space, binomial p-model, $\mathcal{B}_p$ as a restriction of the space of independent binomials, $\mathcal{I}_p$, as follows

$$\mathbb{P}_{\mathcal{B}_p}\big(\boldsymbol{S}=\boldsymbol{s},\boldsymbol{T}=\boldsymbol{t}\big) := \mathbb{P}_{\mathcal{I}_p}\Big(\boldsymbol{S}=\boldsymbol{s},\boldsymbol{T}=\boldsymbol{t} \mid \sum_j S_j = \sum_k T_k\Big).$$

As we alluded to earlier, finding a corresponding binomial model for the graph p-model requires more work than for the other random graph models. This binomial p-model will be used as a stepping stone in constructing a probability model that approximates the graph p-model. Indeed in Definition 7.7 we will introduce a binomial *integrated*-model, $\mathcal{V}_p$, which is a convolution of $\mathcal{B}_p$ with a normal distribution. This binomial *integrated*-model, $\mathcal{V}_p$ is shown to correspond to the graph p-model in Section 7.2.4.

We will calculate the probability of an $(m+n)$-tuple, $(\boldsymbol{s},\boldsymbol{t})$ in the binomial p-model. However, this result is left until Section 7.2.3 as it uses results from the following section where we introduce the binomial *edge*-model.

7.1.3 Binomial *edge*-model, $\mathcal{B}_M$.

The random variables in the binomial *edge*-model, $\mathcal{B}_M$, inherit the property $(\sum_j S_j = \sum_k T_k)$ from the binomial p-model just defined. The parameter M will correspond to the number of edges in the bipartite graph. So in this model we have the extra constraint on $(\boldsymbol{S},\boldsymbol{T})$ which corresponds to the number of edges in the bipartite graph. Hence we require the stricter property $(\sum_j S_j = \sum_k T_k = M)$, i.e. $(\boldsymbol{S},\boldsymbol{T}) \in I_{m,n,M}$.

Definition 7.3 (Binomial *edge*-model, $\mathcal{B}_M$.)**.** The **binomial *edge*-model** $\mathcal{B}_M = \mathcal{B}_M(m,n)$ has support $I_{m,n,M}$. We formally define the probability space, binomial *edge*-model, $\mathcal{B}_M$

as a restriction of the space of independent binomials, $\mathcal{I}_p$, as follows

$$\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{S} = \boldsymbol{s}, \boldsymbol{T} = \boldsymbol{t}) := \mathbb{P}_{\mathcal{I}_p}\Big(\boldsymbol{S} = \boldsymbol{s}, \boldsymbol{T} = \boldsymbol{t} \mid \sum_j S_j = \sum_k T_k = M\Big).$$

We now show an example calculation in the binomial *edge*-model, $\mathcal{B}_M = \mathcal{B}_M(4,3)$ where our parameter, M is equal to six. The probability of the $(4+3)$-tuple, $(1,3,1,1;2,1,3)$ is then,

$$\begin{aligned}
\mathbb{P}_{\mathcal{B}_{M=6}}((1,3,1,1)(2,1,3)) &= \frac{\binom{3}{1}pq^2\binom{3}{3}p^3\binom{3}{1}pq^2\binom{3}{1}pq^2\binom{4}{2}p^2q^2\binom{4}{1}pq^3\binom{4}{3}p^3q}{\binom{12}{6}p^6q^6\binom{12}{6}p^6q^6} \\
&= \frac{\binom{3}{1}\binom{3}{3}\binom{3}{1}\binom{3}{1}\binom{4}{2}\binom{4}{1}\binom{4}{3}}{\binom{12}{6}\binom{12}{6}} = \frac{18}{77^2}.
\end{aligned}$$

Observe that in the calculation above, all appearances of p and q cancelled.

We show this is true in general, i.e. the probability of an $(m+n)$-tuple $(\boldsymbol{s}, \boldsymbol{t})$ in the binomial *edge*-model, $\mathcal{B}_M$, is independent of p. In the following lemma, we calculate the probability of $(\boldsymbol{s}, \boldsymbol{t})$ in the binomial *edge*-model, $\mathcal{B}_M$, where $(\boldsymbol{s}, \boldsymbol{t})$ is an $(m+n)$-tuple that has some of the necessary properties of a degree sequence of a bipartite graph with M edges on vertices (m, n). Hence, we require that $(\boldsymbol{s}, \boldsymbol{t}) \in \{0, 1, \dots, n\}^m \times \{0, 1, \dots, m\}^n$ satisfies $\sum_j s_j = \sum_k t_k = M$, i.e. we require $(\boldsymbol{s}, \boldsymbol{t}) \in I_{m,n,M}$.

Lemma 7.4. *Fix an $(m+n)$-tuple $(\boldsymbol{s}, \boldsymbol{t}) \in I_{m,n,M}$. Then we have*

$$\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{S} = \boldsymbol{s}, \boldsymbol{T} = \boldsymbol{t}) = \binom{mn}{M}^{-2} \prod_{j=1}^{m} \binom{n}{s_j} \prod_{k=1}^{n} \binom{m}{t_k}.$$

Proof. This follows by the short calculation given below. Note the first line is the definition of the binomial *edge*-model, $\mathcal{B}_M$. Then, (7.1) follows by Bayes law. In (7.2) we factor the denominator using the independence of all random variables in the *independent*-model, $\mathcal{I}_p$ and can simplify the numerator because we have already required that $(s,t) \in I_{m,n,M}$.

$$\begin{aligned}
\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{S} = \boldsymbol{s}, \boldsymbol{T} = \boldsymbol{t}) &= \mathbb{P}_{\mathcal{I}_p}\Big(\boldsymbol{S} = \boldsymbol{s}, \boldsymbol{T} = \boldsymbol{t} \;\mid\; \sum_j S_j = \sum_k T_k = M\Big) \\
&= \frac{\mathbb{P}_{\mathcal{I}_p}\Big(\boldsymbol{S} = \boldsymbol{s}, \boldsymbol{T} = \boldsymbol{t} \quad \& \quad \sum_j S_j = \sum_k T_k = M\Big)}{\mathbb{P}_{\mathcal{I}_p}\Big(\sum_j S_j = \sum_k T_k = M\Big)} \qquad (7.1) \\
&= \frac{\mathbb{P}_{\mathcal{I}_p}(\boldsymbol{S} = \boldsymbol{s}, \boldsymbol{T} = \boldsymbol{t})}{\mathbb{P}_{\mathcal{I}_p}\Big(\sum_j S_j = M\Big)\mathbb{P}_{\mathcal{I}_p}\Big(\sum_k T_k = M\Big)} \qquad (7.2)
\end{aligned}$$

In (7.2), $\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{S} = \boldsymbol{s}, \boldsymbol{T} = \boldsymbol{t})$ has been reduced to known quantities. Line (7.3) now follows by the definition of the *independent*-model, $\mathcal{I}_p$. Lastly, in (7.3) all instances of p and q cancel and we have our result.

$$= \prod_{j=1}^{m} \binom{n}{s_j} p^{s_j} q^{n-s_j} \prod_{k=1}^{n} \binom{m}{t_k} p^{t_k} q^{m-t_k} \left(\binom{mn}{M} p^M q^{mn-M}\right)^{-2} \tag{7.3}$$

$$= \binom{mn}{M}^{-2} \prod_{j=1}^{m} \binom{n}{s_j} \prod_{k=1}^{n} \binom{m}{t_k} \tag{7.4}$$

□

Later, Theorem 7.1 shows that for non-*pathological* degree sequences, $(\boldsymbol{s}, \boldsymbol{t})$, $\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{s}, \boldsymbol{t})$ is asymptotically very close to the probability of $(\boldsymbol{s}, \boldsymbol{t})$ occurring in the graph *edge*-model, $\mathcal{G}_M$ under suitable constraints on M.

7.1.4 Binomial *half*-model, $\mathcal{B}_{\boldsymbol{t}}$.

We define a binomial model which will corresponds to the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$. We term this the binomial *half*-model and denote it by $\mathcal{B}_{\boldsymbol{t}}$.

Recall that the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$, is a probability space unique to bipartite graphs. In this graph model, we fix the degrees of our *black* vertices and are interested in the likely distribution of the degrees of the *white* vertices, given this constraint.

The binomial *half*-model, $\mathcal{B}_{\boldsymbol{t}}$, inherits both the property that $(\sum_j S_j = \sum_k T_k)$ from the binomial p-model, $\mathcal{B}_p$, and the stricter property that $(\sum_j S_j = \sum_k T_k = M)$ from the binomial *edge*-model, $\mathcal{B}_M$. Also, in the binomial *half*-model we have the yet stricter constraint that the values of the T_k (and not just their sum) are fixed. That is, we fix the values of the random variables $\boldsymbol{T} = (T_1, T_2, \ldots, T_n)$.

The intuition behind this constraint is that the random variables, $(T_1, T_2, \ldots, T_n)$, in the binomial *half*-model, $\mathcal{B}_{\boldsymbol{t}}$, will correspond to the random variables for the degrees of the *black* vertices, $(T_1, T_2, \ldots, T_n)$, in the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$. In particular, one of the parameters of the binomial *half*-model, $\mathcal{B}_{\boldsymbol{t}}$, is the n-tuple $\boldsymbol{t}$ and we work in the subset of the domain of $\mathcal{I}_p$ where $\boldsymbol{T} = \boldsymbol{t}$. We define this concept formally below.

Definition 7.5 (Binomial *half*-model, $\mathcal{B}_{\boldsymbol{t}}$.)**.** The **binomial *half*-model** $\mathcal{B}_{\boldsymbol{t}} = \mathcal{B}_{\boldsymbol{t}}(m, n, \boldsymbol{t})$ has support $I_{n,m,\boldsymbol{t}}$. We formally define binomial *half*-model, $\mathcal{B}_{\boldsymbol{t}}$, as a restriction of the space $\mathcal{I}_p$. Note $\boldsymbol{t}$ is a parameter of the space.

$$\mathbb{P}_{\mathcal{B}_{\boldsymbol{t}}}(\boldsymbol{S}=\boldsymbol{s},\boldsymbol{T}=\boldsymbol{t}) := \mathbb{P}_{\mathcal{I}_p}\Big(\boldsymbol{S}=\boldsymbol{s},\boldsymbol{T}=\boldsymbol{t} \mid \boldsymbol{T}=\boldsymbol{t}, \sum_j S_j = \sum_k t_k\Big)$$

In the binomial *half*-model, $\mathcal{B}_{\boldsymbol{t}}$, the probability of obtaining an (m+n)$-$*tuple* is independent of the parameter p. We showed this for the *edge*-model in Lemma 7.4. Recall that the binomial *half*-model is defined as a subspace of the binomial *independent*-model, $\mathcal{I}_p$, which is dependent on p. We prove this result in the following lemma.

Lemma 7.6. *Fix an $(m+n)$-tuple $(\boldsymbol{s},\boldsymbol{t}) \in I_{m,n,\boldsymbol{t}}$ and set $M=\sum_k t_k$. Then we have*

$$\mathbb{P}_{\mathcal{B}_{\boldsymbol{t}}}(\boldsymbol{S}=\boldsymbol{s},\boldsymbol{T}=\boldsymbol{t}) = \binom{mn}{M}^{-1}\prod_{j=1}^{m}\binom{n}{s_j}.$$

Proof. The proof is very similar to that for Lemma 7.4.

Note the first line, (7.5) is the definition of the binomial *half*-model, $\mathcal{B}_{\boldsymbol{t}}$. Line (7.6) follows by Bayes law. In (7.7) we factor the denominator using the independence of all random variables in the *independent*-model, $\mathcal{I}_p$ and we can simplify the numerator because we have already required that $(s,t) \in I_{m,n,\boldsymbol{t}}$.

$$\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{S}=\boldsymbol{s},\boldsymbol{T}=\boldsymbol{t}) = \mathbb{P}_{\mathcal{I}_p}\Big(\boldsymbol{S}=\boldsymbol{s},\boldsymbol{T}=\boldsymbol{t} \;\mid\; \sum_j S_j = \sum_k T_k \;\;\&\;\; \boldsymbol{T}=\boldsymbol{t}\Big) \tag{7.5}$$

$$= \frac{\mathbb{P}_{\mathcal{I}_p}\Big(\boldsymbol{S}=\boldsymbol{s} \;\;\&\;\; \boldsymbol{T}=\boldsymbol{t} \;\;\&\;\; \sum_j S_j = \sum_k T_k \;\;\&\;\; \boldsymbol{T}=\boldsymbol{t}\Big)}{\mathbb{P}_{\mathcal{I}_p}\Big(\sum_j S_j = M \;\;\&\;\; \boldsymbol{T}=\boldsymbol{t}\Big)} \tag{7.6}$$

$$= \frac{\mathbb{P}_{\mathcal{I}_p}\big(\boldsymbol{S}=\boldsymbol{s}\big)\mathbb{P}_{\mathcal{I}_p}\big(\boldsymbol{T}=\boldsymbol{t}\big)}{\mathbb{P}_{\mathcal{I}_p}\Big(\sum_j S_j = M\Big)\mathbb{P}_{\mathcal{I}_p}\Big(\boldsymbol{T}=\boldsymbol{t}\Big)} \tag{7.7}$$

We are almost done. Line (7.8) follows by the definition of the *independent*-model, $\mathcal{I}_p$. Lastly, all instances of p and q cancel in (7.9) and we have our result.

$$\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{S}=\boldsymbol{s},\boldsymbol{T}=\boldsymbol{t}) = \prod_{j=1}^{m}\binom{n}{s_j}p^{s_j}q^{n-s_j}\left(\binom{mn}{M}p^M q^{mn-M}\right)^{-1} \tag{7.8}$$

$$= \binom{mn}{M}^{-1}\prod_{j=1}^{m}\binom{n}{s_j} \tag{7.9}$$

□

7.2 Relation to random graph models

In the previous section we defined some binomially based probability spaces. In this section we will show that the probability of a given (non-*pathological*) degree sequence $(\boldsymbol{s}, \boldsymbol{t})$ in each of the random graph models is asymptotically very close to the probability of the same $(m+n)$-tuple $(\boldsymbol{s}, \boldsymbol{t})$ in one of our newly defined binomial models.

Of these correspondences between our three graph models and the binomial models, two of these are easily shown. We do these cases first. In Section 7.2.1 we show that the random graph *edge*-model, $\mathcal{G}_M$, can be approximated by our binomial *edge*-model, $\mathcal{B}_M$. Then in a similar fashion we show that the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$, can be approximated by the binomial *half*-model, $\mathcal{B}_{\boldsymbol{t}}$, in Section 7.2.2. The third random graph model, the p-model, $\mathcal{G}_p$, requires more work. We will construct another probability space, the binomial *integrated*-model, $\mathcal{V}_p$, based on the binomial p-model, $\mathcal{B}_p$. After some calculations we show that this newly defined binomial *integrated*-model, $\mathcal{V}_p$, can be used to approximate the probability of non-*pathological* degree sequences $(\boldsymbol{s}, \boldsymbol{t})$ in the graph p-model, $\mathcal{G}_p$. This final case is done is Section 7.2.3.

7.2.1 Binomial *edge*-model, $\mathcal{B}_M \sim$ graph *edge*-model, $\mathcal{G}_M$.

We show that for non-*pathological*degree sequences $(\boldsymbol{s}, \boldsymbol{t})$ the probability of that degree sequence occurring in the graph *edge*-model, $\mathcal{G}_M$, is very close to the probability of the same $(m+n)$-tuple in the binomial *edge*-model, $\mathcal{B}_M$.

Theorem 7.1. *Fix $a, b \in \mathbb{R}^+$ such that $a+b < \frac{1}{2}$. Let $\mathcal{G}_M = \mathcal{G}_M(m,n)$ be as in Definition 1.7 and $\mathcal{B}_M = \mathcal{B}_M(m,n)$ as in Definition 7.3. Then there exists $\varepsilon > 0$ such that as $m, n \to \infty$ subject to Conditions 5.1, for non-(a,b,m,n,ε)-*pathological *degree sequence $(\boldsymbol{s}, \boldsymbol{t}) \in I_{m,n,M}$,*

$$\mathbb{P}_{\mathcal{G}_M}(\boldsymbol{s}, \boldsymbol{t}) \;\;=\;\; \mathbb{P}_{\mathcal{B}_M}(\boldsymbol{s}, \boldsymbol{t})(1 + O(n^{-b})).$$

Proof. We first calculate the probability of a fixed non-*pathological* degree sequence $(\boldsymbol{s}, \boldsymbol{t})$ in the graph *edge*-model, $\mathcal{G}_M$. From Definition 1.6 we have

$$\mathbb{P}_{\mathcal{G}_M}(\boldsymbol{s}, \boldsymbol{t}) := \mathbb{P}_{\mathcal{G}_M}(H : \textit{the degree sequence of } H \textit{ is } (\boldsymbol{s}, \boldsymbol{t})).$$

Hence $\mathbb{P}_{\mathcal{G}_M}(\boldsymbol{s}, \boldsymbol{t})$ can be calculated by taking the number of bipartite graphs with degree sequence $(\boldsymbol{s}, \boldsymbol{t})$ and dividing by the total number of labelled bipartite graphs on (m, n) vertices with M edges. There are $\binom{mn}{M}$ graphs on (m, n) vertices with M edges. Let us denote the number of bipartite graphs with degree sequence $(\boldsymbol{s}, \boldsymbol{t})$ by $|B(\boldsymbol{s}, \boldsymbol{t})|$. Hence,

$$\mathbb{P}_{\mathcal{G}_M}(\boldsymbol{s},\boldsymbol{t}) := \binom{mn}{M}^{-1} |B(\boldsymbol{s},\boldsymbol{t})|. \tag{7.10}$$

By Corollary 2.10, for p which is *acceptable* for a, m and n and non-(a,b,m,n,ε)-*pathological* degree sequences, $(\boldsymbol{s},\boldsymbol{t})$, we have an approximate count of the number of graphs with that degree sequence, i.e. an approximate value for $|B(\boldsymbol{s},\boldsymbol{t})|$. We substitute this value for $|B(\boldsymbol{s},\boldsymbol{t})|$ in (7.10) to yield,

$$\mathbb{P}_{\mathcal{G}_M}(\boldsymbol{s},\boldsymbol{t}) \quad = \quad \binom{mn}{M}^{-2} \prod_{j=1}^{m} \binom{n}{s_j} \prod_{k=1}^{n} \binom{m}{t_k} \exp(O(n^{-b})). \tag{7.11}$$

The form of the right-hand side is very similar to that for the probability of an $(m+n)$-tuple $(\boldsymbol{s},\boldsymbol{t})$ in the binomial *edge*-model $\mathcal{B}_M$. In particular by Lemma 7.4,

$$\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{s},\boldsymbol{t}) = \binom{mn}{M}^{-2} \prod_{j=1}^{m} \binom{n}{s_j} \prod_{k=1}^{n} \binom{m}{t_k}. \tag{7.12}$$

We now substitute the expression for $\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{s},\boldsymbol{t})$ in (7.12) into line (7.11) to yield the required result.

□

7.2.2 Binomial *half*-model, $\mathcal{B}_{\boldsymbol{t}}$ ∼ graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$.

In this section we show that the probability of an (ε, a)-*regular* degree sequence $(\boldsymbol{s},\boldsymbol{t})$ in the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$, is asymptotically very close to the probability of $(\boldsymbol{s},\boldsymbol{t})$ in the binomial *half*-model, $\mathcal{B}_{\boldsymbol{t}}$.

This proceeds by a direct proof which compares results for the probability of $(m+n)$-tuples $(\boldsymbol{s},\boldsymbol{t})$ in the binomial *half*-model, $\mathcal{B}_{\boldsymbol{t}}$, to results for the probability of (ε, a)-*regular* degree sequences $(\boldsymbol{s},\boldsymbol{t})$ in the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$.

Theorem 7.2. *Let $\mathcal{G}_{\boldsymbol{t}} = \mathcal{G}_{\boldsymbol{t}}(m,n)$ be as in Definition 1.8 and $\mathcal{B}_{\boldsymbol{t}} = \mathcal{B}_{\boldsymbol{t}}(m,n)$ be as in Definition 7.5. Fix a non-*pathological *degree sequence $(\boldsymbol{s}(m,n), \boldsymbol{t}(m,n)) \in I_{m,n,\boldsymbol{t}}$. Then as $m, n \to \infty$ subject to Conditions 6.1,*

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}\big(\boldsymbol{s},\boldsymbol{t}\big) \quad = \quad \mathbb{P}_{\mathcal{B}_{\boldsymbol{t}}}\big(\boldsymbol{s},\boldsymbol{t}\big)\big(1 + O(n^{-b})\big).$$

Proof. We first calculate the probability of a fixed non-*pathological* degree sequence $(\boldsymbol{s},\boldsymbol{t})$ in the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$. From Definition 1.8 of the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$ we have

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(\boldsymbol{s},\boldsymbol{t}) := \mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(H : \textit{the degree sequence of } H \textit{ is } (\boldsymbol{s},\boldsymbol{t})).$$

Hence $\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(\boldsymbol{s},\boldsymbol{t})$ can be calculated by taking the number of bipartite graphs with degree sequence $(\boldsymbol{s},\boldsymbol{t})$ and dividing by the total number of labelled bipartite graphs on (m,n) vertices with *black* degree sequence $\boldsymbol{t}$. There are $\prod_k \binom{m}{t_k}$ graphs on (m,n) vertices with *black* degree sequence $\boldsymbol{t}$[1]. So we have

$$\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(\boldsymbol{s},\boldsymbol{t}) := \left(\prod_{k=1}^{n}\binom{m}{t_k}\right)^{-1} |B(\boldsymbol{s},\boldsymbol{t})|. \tag{7.13}$$

By Corollary 2.10 from [GM09], for *acceptable* for a, m and n values of the n-tuple $\boldsymbol{t}$, and non-(a,b,m,n,ε)-*pathological* degree sequences, $(\boldsymbol{s},\boldsymbol{t})$, we have an approximate count for the number of graphs with that degree sequence, i.e. an approximate value of $|B(\boldsymbol{s},\boldsymbol{t})|$. We substitute this value for $|B(\boldsymbol{s},\boldsymbol{t})|$ in (7.13) to yield,

$$\begin{aligned}\mathbb{P}_{\mathcal{G}_{\boldsymbol{t}}}(\boldsymbol{s},\boldsymbol{t}) &= \left(\binom{mn}{M}\prod_{k=1}^{n}\binom{m}{t_k}\right)^{-1}\prod_{j=1}^{m}\binom{n}{s_j}\prod_{k=1}^{n}\binom{m}{t_k}\exp(O(n^{-b}))\\ &= \binom{mn}{M}^{-1}\prod_{j=1}^{m}\binom{n}{s_j}\exp(O(n^{-b})). \end{aligned}\tag{7.14}$$

The form of the right-hand side is very similar to that for the probability of an $(m+n)$-tuple $(\boldsymbol{s},\boldsymbol{t})$ in the binomial *half*-model, $\mathcal{B}_{\boldsymbol{t}}$. In particular by Lemma 7.6, $\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{s},\boldsymbol{t}) = \binom{mn}{M}^{-1}\prod_{j=1}^{m}\binom{n}{S_j}$. In light of this we may substitute $\mathbb{P}_{\mathcal{B}_{\boldsymbol{t}}}(\boldsymbol{s},\boldsymbol{t})$ into the equation (7.14) above to give the required result.

□

[1]To see this note the following. Fix a *black* vertex v_k. There are t_k edges incident with this vertex. Hence of the m *white* vertices, there are edge connecting v_k to any t_k distinct *white* vertices. So, there are $\binom{m}{t_k}$ possible arrangements for the edges incident with v_k. Now, note the edges incident with one *black* vertex are disjoint from the edges incident with any other *black* vertex. Hence the total number of labelled bipartite graphs with *black* degree sequence $\boldsymbol{t}$ is the product: $\prod_{k=1}^{n}\binom{m}{t_k}$.

7.2.3 Binomial p-model, $\mathcal{B}_p$.

Given an $(m+n)$-tuple $(\boldsymbol{s},\boldsymbol{t})$ we will calculate the probability of $(\boldsymbol{s},\boldsymbol{t})$ in the binomial p-model, $\mathcal{B}_p$ but first we need the following somewhat technical lemma.

Lemma 7.3. *Let $0 < a < 1/2$ and suppose that $pq > \frac{1}{\log n}$. Then,*

$$\sum_{i=1}^{n} \binom{n}{i} p^i q^{n-i} = \frac{1}{2\sqrt{\pi npq}} \left(1 + O\left(n^{-1/2+\varepsilon}\right)\right) + O(e^{-n^{\varepsilon}})$$

Corollary 7.4. *Let $0 < a < 1/2$ and suppose that p is* acceptable *for a, m and n. Then,*

$$\mathbb{P}_{\mathcal{I}_p}\big(\sum_j S_j = \sum_k T_k\big) = \frac{1}{2\sqrt{\pi pqmn}}\big(1 + O(n^{-1})\big) + O(e^{-n^{\varepsilon}})$$

Proof. Each S_j is binomially distributed in $\mathcal{I}_p$ with parameters (n, p). Thus $\sum_{j=1}^{m} S_j$ is binomially distributed with parameters (mn, p). Similarly, $\sum_{k=1}^{n} T_k$ is also binomially distributed in $\mathcal{I}_p$ with parameters (mn, p).

Thus we have,

$$\mathbb{P}_{\mathcal{I}_p}\big(\sum_j S_j = \sum_k T_k\big) = \sum_{i=1}^{mn} \binom{mn}{i} p^i q^{mn-i}.$$

Hence it is sufficient to prove that,

$$\sum_{i=1}^{mn} \binom{mn}{i} p^i q^{mn-i} = \frac{1}{2\sqrt{\pi pqmn}}\big(1 + O(n^{-1})\big) + O(e^{-n^{\varepsilon}})$$

and so the result follows by Lemma 7.3 □

Lemma 7.5. *Fix $(\boldsymbol{s},\boldsymbol{t}) \in I_{m,n}$ and set $M = \sum_j s_j$. Then let $\mathcal{B}_p = \mathcal{B}_p(m,n)$ be as in Definition 7.2 and $\mathcal{B}_M = \mathcal{B}_M(m,n,M)$ be as in Definition 7.3. Then as $m, n \to \infty$ subject to Conditions 4.1,*

$$\mathbb{P}_{\mathcal{B}_p}(\boldsymbol{s},\boldsymbol{t}) = \prod_{j=1}^{m} \binom{n}{s_j} \prod_{k=1}^{n} \binom{m}{t_k} p^{2M} q^{2mn-2M} 2\sqrt{\pi pqmn}\big(1 + O(n^{-1/2} \ln^2 n)\big).$$

Proof. By Bayes law,

$$\mathbb{P}_{\mathcal{B}_p}(\boldsymbol{s},\boldsymbol{t}) = \mathbb{P}_{\mathcal{B}_p}\Big(\sum_j S_j = M\Big)\ \mathbb{P}_{\mathcal{B}_p}\Big((\boldsymbol{s},\boldsymbol{t}) \mid \sum_j S_j = \sum_k T_k = M\Big).$$

Then,

$$\begin{aligned}\mathbb{P}_{\mathcal{B}_p}(\boldsymbol{s},\boldsymbol{t}) &= \mathbb{P}_{\mathcal{I}_p}\Big(\sum_j S_j = \sum_k T_k = M \ \mid\ \sum_j S_j = \sum_k T_k\Big)\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{s},\boldsymbol{t})\\ &= \frac{\mathbb{P}_{\mathcal{I}_p}\big(\sum_j S_j = \sum_k T_k = M \ \&\ \sum_j S_j = \sum_k T_k\big)}{\mathbb{P}_{\mathcal{I}_p}(\sum_j S_j = \sum_k T_k)}\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{s},\boldsymbol{t})\end{aligned}$$

where the first line follows by the definitions of $\mathcal{B}_p$ and $\mathcal{B}_M$ and second line follows again by Bayes Law. Observe now that the two statements,

$$\sum_j S_j = \sum_k T_k = M \ \&\ \sum_j S_j = \sum_k T_k$$

and

$$\sum_j S_j = M \ \&\ \sum_k T_k = M$$

describe the same event. Also recall that in the $\mathcal{I}_p$ model, the S_j's and the T_k's are independent. Hence,

$$\mathbb{P}_{\mathcal{B}_p}(\boldsymbol{s},\boldsymbol{t}) = \frac{1}{\mathbb{P}_{\mathcal{I}_p}(\sum_j S_j = \sum_k T_k)}\mathbb{P}_{\mathcal{I}_p}\big(\sum_j S_j = M\big)\mathbb{P}_{\mathcal{I}_p}\big(\sum_k T_k = M\big)\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{s},\boldsymbol{t}). \qquad (7.15)$$

The values of each of the terms on the right hand side of (7.15) are known. The quantities, $\mathbb{P}_{\mathcal{I}_p}\big(\sum_j S_j = M\big)$ and $\mathbb{P}_{\mathcal{I}_p}\big(\sum_k T_k = M\big)$ follow straight from Definition 7.1 of the $\mathcal{I}_p$ model. The calculations for $\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{s},\boldsymbol{t})$ appear in Lemma 7.4. Substitute these values into (7.15) to yield,

$$\mathbb{P}_{\mathcal{B}_p}(\boldsymbol{s},\boldsymbol{t}) = \frac{1}{\mathbb{P}_{\mathcal{I}_p}\Big(\sum_j S_j = \sum_k T_k\Big)}(\binom{mn}{M}p^M q^{mn-M})^2\binom{mn}{M}^{-2}\prod_{j=1}^{m}\binom{n}{s_j}\prod_{k=1}^{n}\binom{m}{t_k}.$$

Lemma 7.4 provides an approximation for $\mathbb{P}_{\mathcal{I}_p}(\sum_j S_j = \sum_k T_k)$. Hence we now have,

$$\mathbb{P}_{\mathcal{B}_p}(\boldsymbol{s},\boldsymbol{t}) = 2\sqrt{\pi pqmn}\big(p^M q^{mn-M}\big)^2\prod_{j=1}^{m}\binom{n}{s_j}\prod_{k=1}^{n}\binom{m}{t_k}\big(1+O(n^{-1/2}\ln^2 n), \qquad (7.16)$$

which completes the proof.

□

By Theorem 4.9 we have a formula to enumerate all bipartite graphs with any given non-*pathological* degree sequence. We use this below to calculate the asymptotic probability of a graph having degree sequence $(\boldsymbol{s},\boldsymbol{t})$.

Lemma 7.6. *Fix $a, b \in \mathbb{R}^+$ and $\varepsilon > 0$ such that $0 < a + b < \frac{1}{2}$. Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.6. Fix non-(a,b,m,n,ε)*-pathological $(\boldsymbol{s},\boldsymbol{t}) \in I_{m,n}$. *Then as $m, n \to \infty$ subject to Conditions 4.1,*

$$\mathbb{P}_{\mathcal{G}_p}(\boldsymbol{s},\boldsymbol{t}) = p^M q^{mn-M} \binom{mn}{M}^{-1} \prod_{j=1}^{m} \binom{n}{s_j} \prod_{k=1}^{n} \binom{m}{t_k} \big(1 + O(n^{-b})\big).$$

Proof. By Bayes law and by the definitions of $\mathcal{G}_p$ and $\mathcal{G}_M$ we have,

$$\begin{aligned}\mathbb{P}_{\mathcal{G}_p}(\boldsymbol{s},\boldsymbol{t}) &= \mathbb{P}_{\mathcal{G}_p}\big(\sum_j S_j = M\big)\, \mathbb{P}_{\mathcal{G}_p}\big((\boldsymbol{s},\boldsymbol{t}) \mid \sum_j S_j = M\big) \\ &= \mathbb{P}_{\mathcal{G}_p}\Big(\sum_j S_j = M\Big)\, \mathbb{P}_{\mathcal{G}_M}\big(\boldsymbol{s},\boldsymbol{t}\big).\end{aligned}$$

Note the probability that a random graph in $\mathcal{G}_p$ has M edges is $\binom{mn}{M} p^M q^{mn-M}$. Also, recall that by line (7.10) we have a formula for $\mathbb{P}_{\mathcal{G}_M}\big(\boldsymbol{s},\boldsymbol{t}\big)$ in terms of the number of bipartite graphs on (m,n) vertices, $|B(\boldsymbol{s},\boldsymbol{t})|$. Thus,

$$\begin{aligned}\mathbb{P}_{\mathcal{G}_p}(\boldsymbol{s},\boldsymbol{t}) &= \binom{mn}{M} p^M q^{mn-M} \binom{mn}{M}^{-1} |B(\boldsymbol{s},\boldsymbol{t})| \\ &= p^M q^{mn-M} \binom{mn}{M}^{-1} \prod_{j=1}^{m} \binom{n}{s_j} \prod_{k=1}^{n} \binom{m}{t_k} \big(1 + O(n^{-b})\big),\end{aligned}$$

where the second line follows by the result of Greenhill and Mckay, Corollary 2.10. $\square$

7.2.4 Integrated binomial model, $\mathcal{V}_p \sim$ graph p-model $\mathcal{G}_p$.

The binomial *integrated*-model, $\mathcal{V}_p$ is defined in terms of the binomial p-model, $\mathcal{B}_M$. We construct $\mathcal{V}_p$ as a convolution of the binomial p-model, $\mathcal{B}_p$, with a normal distribution, $\mathcal{K}_p$.

Currently the binomial p-model, $\mathcal{B}_p$, is defined only for $p \in [0,1]$, so we let $\mathbb{P}_{\mathcal{B}_p}(\boldsymbol{s},\boldsymbol{t}) := 0, \quad \forall p \notin [0,1]$. In the definition below we normalise the probability space by dividing through by $V(p)$.

Definition 7.7. *The **binomial** integrated-**model** $\mathcal{V}_p = \mathcal{V}_p(m,n,p)$ has support $I_{n,m}$. Let $\mathcal{K}_p = \left(\frac{mn}{\pi pq}\right)^{1/2} \exp\left(\frac{-mn}{pq}(p-p')^2\right)$ and $V(p) = \int_0^1 \mathcal{K}_p(p')dp'$. Then define,*

$$\mathbb{P}_{\mathcal{V}_p}(\boldsymbol{s},\boldsymbol{t}) := \frac{1}{V(p)} \int_{-\infty}^{\infty} \mathcal{K}_p(p')\mathbb{P}_{\mathcal{B}_{p'}}(\boldsymbol{s},\boldsymbol{t})dp'.$$

In this section we will calculate the probability of a given degree sequence $(\boldsymbol{s},\boldsymbol{t})$ in the model $\mathcal{V}_p$ and show that this agrees very closely with the probability of this degree sequence occurring in our random graph p-model, $\mathcal{G}_p$. This result will rely on the following two technical lemmas.

Lemma 7.8. *Let $y := \frac{M-pmn}{\sqrt{pqmn}}$ and assume that $|y| < n^{4\varepsilon}$. Then as $m,n \to \infty$ subject to p* acceptable *for a,m and n,*

$$\left(p^M q^{mn-M}\binom{mn}{M}\right)^{-1} = \sqrt{2\pi pqmn}\exp\left(\frac{y^2}{2} + O\big((mn)^{-1/2+\varepsilon}\big)\right)$$

Lemma 7.9. *Fix $\varepsilon > 0$. Let $y = \frac{M-pmn}{\sqrt{pqmn}}$ and $\delta = p - p'$. Write $q = 1-p$ and $q' = 1-p'$. Assume $|y| < n^{4\varepsilon}$. Then as $m,n \to \infty$ subject to p* acceptable *for a,m and n,*

$$\int_{-\infty}^{\infty} \left(\frac{p'}{p}\right)^{2M+1/2}\left(\frac{q'}{q}\right)^{2mn-2M+1/2} \exp\left(-\frac{mn}{pq}\delta^2\right)d\delta = \sqrt{\frac{\pi pq}{2mn}}\exp\left(\frac{y^2}{2}\right)\left(1+O(n^{-1/2+12\varepsilon})\right)$$

We defer the proof of both Lemma 7.8 and 7.9 to Section 9.1 of the Appendix. Assuming Lemmas 7.8 and 7.9 we are now in a position to show that for non-*pathological* $(\boldsymbol{s},\boldsymbol{t})$, $\mathbb{P}_{\mathcal{V}_p}(\boldsymbol{s},\boldsymbol{t})$ is a good asymptotic approximation for $\mathbb{P}_{\mathcal{G}_p}(\boldsymbol{s},\boldsymbol{t})$.

Theorem 7.10. *Fix $a,b \in \mathbb{R}^+$ and $\varepsilon < \varepsilon_0(a,b)$ such that $a+b < \frac{1}{2}$, $a = a' + \varepsilon < \frac{1}{2}$ and $b + 17\varepsilon < 1$. Let $\mathcal{G}_p = \mathcal{G}_p(m,n)$ be as in Definition 1.6 and $\mathcal{V}_p = \mathcal{V}_p(m,n)$ be as in Definition 7.7. Let $(\boldsymbol{s},\boldsymbol{t}) \in I_{n,m,M}$ for some $0 \le M \le mn$. Suppose also that $(\boldsymbol{s},\boldsymbol{t})$ is non-(a,b,m,n,ε)-*pathological *and satisfies $\frac{M-pmn}{\sqrt{pqmn}} < n^{5\varepsilon}$.*
Then as $m,n \to \infty$ subject to Conditions 4.1,

$$\mathbb{P}_{\mathcal{G}_p}(\boldsymbol{s},\boldsymbol{t}) = \mathbb{P}_{\mathcal{V}_p}(\boldsymbol{s},\boldsymbol{t})\big(1+O(n^{-b})\big).$$

Proof.

$$\begin{aligned}
&\mathbb{P}_{\mathcal{V}_p}(\boldsymbol{s},\boldsymbol{t})\\
&\quad=\int_{-\infty}^{\infty}\big(\frac{mn}{\pi pq}\big)^{1/2}\exp\big(\frac{-mn}{pq}\delta^2\big)\mathbb{P}_{\mathcal{B}_{p'}}(\boldsymbol{s},\boldsymbol{t})d\delta && (7.17)\\
&\quad=\int_{-\infty}^{\infty}\big(\frac{mn}{\pi pq}\big)^{1/2}\exp\big(\frac{-mn}{pq}\delta^2\big)\prod_{j=1}^{m}\binom{n}{s_j}\prod_{k=1}^{n}\binom{m}{t_k}\\
&\qquad\times p'^{2M}q'^{2mn-2M}2\sqrt{\pi p'q'mn}\big(1+O(n^{-1/2}\ln^2 n)\big)d\delta && (7.18)\\
&\quad=p^Mq^{mn-M}\binom{mn}{M}^{-1}\prod_{j=1}^{m}\binom{n}{s_j}\prod_{k=1}^{n}\binom{m}{t_k}2mnp^Mq^{mn-M}\binom{mn}{M}\big(1+O(n^{-1/2}\ln^2 n)\big)\\
&\qquad\times\int_{-\infty}^{\infty}\exp\big(\frac{-mn}{pq}\delta^2\big)\Big(\frac{p'}{p}\Big)^{2M+1/2}\Big(\frac{q'}{q}\Big)^{2mn-2M+1/2}d\delta && (7.19)\\
&\quad=\mathbb{P}_{\mathcal{G}_p}(\boldsymbol{s},\boldsymbol{t})2mnp^Mq^{mn-M}\binom{mn}{M}\big(1+O(n^{-1/2}\ln^2 n)\big)\big(1+O(n^{-b})\big)\\
&\qquad\times\int_{-\infty}^{\infty}\exp\big(\frac{-mn}{pq}\delta^2\big)\Big(\frac{p'}{p}\Big)^{2M+1/2}\Big(\frac{q'}{q}\Big)^{2mn-2M+1/2}d\delta && (7.20)
\end{aligned}$$

The first line, (7.17), follows by Definition 7.7 of the binomial *integrated*-model, $\mathcal{V}_p$. Then (7.18) follows by Lemma 7.5. Line (7.19) follows by rearranging and cancelling in (7.18). By Lemma 7.6 and because $(\boldsymbol{s},\boldsymbol{t})$ is non-*pathological* we then get (7.20).

Let $y=\frac{M-pmn}{\sqrt{pqmn}}$ and then by assumption $|y|<n^{4\varepsilon}$. Now (7.21) follows by Lemma 7.9 (also note that the $O(n^{-b})$ error term is larger). Hence,

$$
\begin{aligned}
&\mathbb{P}_{\mathcal{V}_p}(\boldsymbol{s},\boldsymbol{t}) \\
&\qquad = \mathbb{P}_{\mathcal{G}_p}(\boldsymbol{s},\boldsymbol{t}) 2mnp^M q^{mn-M}\binom{mn}{M}\big(1+O(n^{-1/2}\ln^2 n)\big)\big(1+O(n^{-b})\big) \\
&\qquad\quad \times \sqrt{\frac{\pi pq}{2mn}}\exp\Big(\frac{y^2}{2}\Big)
\end{aligned} \tag{7.21}
$$

$$
\begin{aligned}
&\qquad = \mathbb{P}_{\mathcal{G}_p}(\boldsymbol{s},\boldsymbol{t}) 2mn\Big(\sqrt{2\pi pqmn}\exp\big(\frac{y^2}{2}+O\big((mn)^{-1/2+\varepsilon}\big)\big)\Big)^{-1} \\
&\qquad\quad \times \big(1+O(n^{-1/2}\ln^2 n)\big)\big(1+O(n^{-b})\big)\sqrt{\frac{\pi pq}{2mn}}\exp\Big(\frac{y^2}{2}\Big)
\end{aligned} \tag{7.22}
$$

$$
\qquad = \mathbb{P}_{\mathcal{G}_p}(\boldsymbol{s},\boldsymbol{t})\Big(1+O\big((mn)^{-1/2+\varepsilon}\big)\Big)\Big(1+O(n^{-1/2}\ln^2 n)\Big)\Big(1+O(n^{-b})\Big) \tag{7.23}
$$

$$
\qquad = \mathbb{P}_{\mathcal{G}_p}(\boldsymbol{s},\boldsymbol{t})\Big(1+O(n^{-b})\Big) \tag{7.24}
$$

Line (7.22) follows by Lemma 7.8. Rearrange to get (7.23). Then because $O(n^{-b})$ is the largest error term we have (7.24) and we are done. □

7.3 Implications

We will show that the expectation of a random variable $X : I_{m,n} \to \mathbb{R}$ in any one of the random bipartite graph models can be asymptotically approximated by the expectation of the same event in the corresponding binomial model. This is done for the graph *edge*-model, $\mathcal{G}_M$ in Theorem 7.4 and the graph p-model, $\mathcal{G}_p$ in Theorem 7.5. The results in Theorems 7.4 and 7.5 form the bipartite analogue of Theorem 2.6 in [MW97], our Theorem 2.6.

First we need some lemmas which bound the probability of pathological $(m+n)$-tuples in each of the binomial models. We begin by considering the binomial p-model, $\mathcal{B}_p$.

Lemma 7.1. *Fix $\varepsilon > 0$ and $0 < a < \frac{1}{2}$ such that $a + \varepsilon < \frac{1}{2}$ and $b + 17\varepsilon < 1$. Let $\mathcal{B}_p = \mathcal{B}_p(m,n)$ be as in Definition 7.2. Then as $m, n \to \infty$ subject to Conditions 4.1,*

$$\mathbb{P}_{\mathcal{B}_p}\big(\ (\boldsymbol{S},\boldsymbol{T})\ \text{ is } (a,b,m,n,\varepsilon)\text{-}\textit{pathological}\ \big) \le e^{-n^{11\varepsilon/10}}.$$

Proof. Many of the bounds we will show in $\mathcal{B}_p$ will follow directly from results in $\mathcal{G}_p$. This is because, considered alone, the degrees of the *white* vertices in $\mathcal{G}_p$ are binomially distributed with parameters (n,p). Also the degrees of u_j and $u_{j'}$ are independent $\forall 0 \le j, j' \le k$ (except $j = j'$). Thus the distribution of the degree of u_j in $\mathcal{G}_p$ is the same as that for the random variables S_j in $\mathcal{B}_p$. (The dependence between the degrees of the vertices u_j and v_k in $\mathcal{G}_p$ is different to the dependence between S_j and T_k is different in $\mathcal{B}_p$.)

Hence by Lemma 4.3,

$$\mathbb{P}_{\mathcal{B}_p}\big(|S_j - S| \ge n^{1/2+2\varepsilon/5}\big) = \mathbb{P}_{\mathcal{G}_p}\big(|S_j - S| \ge n^{1/2+2\varepsilon/5}\big) \le e^{-n^{4\varepsilon/3}},$$

$$\mathbb{P}_{\mathcal{B}_p}\big(|T_k - T| \ge n^{1/2+10\varepsilon/11}\big) = \mathbb{P}_{\mathcal{G}_p}\big(|T_k - T| \ge n^{1/2+10\varepsilon/11}\big) \le e^{-n^{5\varepsilon/4}}.$$

Thus,

$$\begin{aligned}
&\mathbb{P}_{\mathcal{B}_p}\big(\forall j,k, \quad T_k - T,\ \ S_j - S \text{ uniformly } o(n^{1/2+\varepsilon})\big) \\
&\qquad \ge 1 - \mathbb{P}_{\mathcal{G}_p}\big(|S_j - S| \ge n^{1/2+2\varepsilon/5}\big) - \mathbb{P}_{\mathcal{G}_p}\big(|T_k - T| \ge n^{1/2+10\varepsilon/11}\big) \\
&\qquad \ge 1 - e^{-n^{6\varepsilon/5}}.
\end{aligned} \tag{7.25}$$

Define $\lambda = \frac{1}{mn}\sum_j S_j$. Then the distribution of λ in $\mathcal{B}_p$ and the edge density λ in $\mathcal{G}_p$ are the same.

Thus by Lemma 4.7,

$$\mathbb{P}_{\mathcal{B}_p}\big(\lambda \text{ is } \textit{acceptable} \text{ for } a, m \text{ and } n\big) = \mathbb{P}_{\mathcal{G}_p}\big(\lambda \text{ is } \textit{acceptable} \text{ for } a, m \text{ and } n\big) \ge 1 - e^{-n^{3\varepsilon/2}}. \tag{7.26}$$

We can apply the same comparisons with the $\mathcal{G}_p$ method to deduce (by Lemmas 4.6 and 4.8 respectively),

$$\mathbb{P}_{\mathcal{B}_p}\left(\left|1 - \frac{\sum_j (S_j - S)^2}{\lambda(1-\lambda)mn}\right| \ \geq \ n^{-1/2+8\varepsilon} \right) \ \leq \ e^{-n^{6\varepsilon/5}}$$

$$\text{and} \qquad \mathbb{P}_{\mathcal{B}_p}\left(\left|1 - \frac{\sum_k (T_k - T)^2}{\lambda(1-\lambda)mn}\right| \ \geq \ n^{-1/2+9\varepsilon} \right) \ \leq \ e^{-n^{8\varepsilon/7}}.$$

And thus,

$$\mathbb{P}_{\mathcal{B}_p}\left(\left|\Big(1 - \frac{\sum_j (S_j - S)^2}{\lambda(1-\lambda)mn}\Big)\Big(1 - \frac{\sum_k (T_k - T)^2}{\lambda(1-\lambda)mn}\Big)\right| \ \geq \ n^{-1+17\varepsilon} \right) \ \leq \ e^{-n^{9\varepsilon/8}}. \tag{7.27}$$

Hence by (7.25),(7.26) and (7.27),

$$\mathbb{P}_{\mathcal{B}_p}\big(\ (\boldsymbol{S}, \boldsymbol{T}) \ \ \textit{is } (a,b,m,n,\varepsilon)\text{-pathological}\ \big) \leq e^{-n^{6\varepsilon/5}} + e^{-n^{3\varepsilon/2}} + e^{-n^{9\varepsilon/8}} \leq e^{-n^{11\varepsilon/10}}. \quad \square$$

Next we work in the binomial *edge*-model, $\mathcal{B}_M$, and bound the probability that a random $(m+n)$-tuple in $\mathcal{B}_M$ will be *pathological*.

Corollary 7.2. *Fix $\varepsilon > 0$ and $0 < a < \frac{1}{2}$ such that $a + \varepsilon < \frac{1}{2}$ and $b + 17\varepsilon < 1$. Let $\mathcal{B}_M = \mathcal{B}_M(m,n)$ be as in Definition 7.3. Then as $m, n \to \infty$ subject to Conditions 5.1,*

$$\mathbb{P}_{\mathcal{B}_M}\big(\ (\boldsymbol{S}, \boldsymbol{T}) \ \ \text{is } (a,b,m,n,\varepsilon)\textit{-pathological}\ \big) \leq e^{-n^{12\varepsilon/11}}.$$

Proof. Similar to Lemma 5.1 we can conclude that for any event $A \subset I_{m,n,M}$,

$$\mathbb{P}_{\mathcal{B}_M}(A) \leq mn\mathbb{P}_{\mathcal{B}_{p=\frac{M}{mn}}}(A). \tag{7.28}$$

Observe that as we have assumed Conditions 5.1, $\frac{M}{mn}$ is *acceptable* for a, m and n. Thus our choice of $p = \frac{M}{mn}$ implies that p is *acceptable* for a, m and n and we have Conditions 4.1.

Let $P \subset I_{m,n,M}$ be the set of (a,b,m,n,ε)-*pathological* $(m+n)$-tuples in $I_{m,n,M}$. By Lemma 7.1, $\mathbb{P}_{\mathcal{B}_{p=\frac{M}{mn}}}(P)e^{-n^{11\varepsilon/10}}$. Now let the event A in (7.28) be our set P and this yields the asserted result. $\square$

Lastly we consider the binomial *integrated*-model, $\mathcal{V}_p$ and bound the probability of the set of *pathological* $(m+n)$-tuples in $\mathcal{V}_p$.

Corollary 7.3. *Fix $\varepsilon > 0$ and $0 < a < \frac{1}{2}$ such that $a + \varepsilon < \frac{1}{2}$ and $b + 17\varepsilon < 1$. Let $\mathcal{V}_p = \mathcal{V}_p(m, n)$ be as in Definition 7.7. Then as $m, n \to \infty$ subject to Conditions 4.1,*

$$\mathbb{P}_{\mathcal{B}_M}\big(\ (\boldsymbol{S}, \boldsymbol{T})\ \text{ is } (a, b, m, n, \varepsilon)\text{-}\textit{pathological}\ \big) \leq e^{-n^{\varepsilon}}.$$

Proof. Let $P \subset I_{m,n}$ be the subset of $(a, b, m, n, \varepsilon)$-*pathological* $(m+n)$-tuples in $I_{m,n}$. By the definition of $\mathcal{V}_p$ and the change of variables $\delta = p' - p$ we get,

$$\begin{aligned} V(p).\mathbb{P}_{\mathcal{V}_p}\big(P\big) =& \sqrt{\frac{mn}{\pi pq}} \int_0^1 \exp\Big(\frac{-mn}{pq}(p'-p)^2\Big)\mathbb{P}_{\mathcal{B}_{p'}}\big(P\big)dp' \\ =& \sqrt{\frac{mn}{\pi pq}} \int_{[-n^{\varepsilon-1},\ n^{\varepsilon-1}]} \exp\Big(\frac{-mn}{pq}\delta^2\Big)\mathbb{P}_{\mathcal{B}_{p'}}\big(P\big)d\delta \\ &+ \sqrt{\frac{mn}{\pi pq}} \int_{[-p,\ q]-[-n^{\varepsilon-1},\ n^{\varepsilon-1}]} \exp\Big(\frac{-mn}{pq}\delta^2\Big)\mathbb{P}_{\mathcal{B}_{p'}}\big(P\big)d\delta \end{aligned} \tag{7.29}$$

We will bound both terms on the right hand side of (7.29) seperately. The integral over the interval $[-\varepsilon \ln n,\ \varepsilon \ln n]$ is bounded first.

Let $a' = a + \varepsilon$ and note that $a' < 1/2$. Now observe that by Definition 2.5 of acceptable that if p is *acceptable* for a, m and n then $p + n^{\varepsilon-1}$ is *acceptable* for a', m and n. Hence for $p' \in [p - n^{\varepsilon-1}, p + n^{\varepsilon-1}]$, by Lemma 7.1, $\mathbb{P}_{\mathcal{B}_{p'}}\big(P\big) \leq e^{-n^{11\varepsilon/10}}$. Thus,

$$\int_{[-n^{\varepsilon-1},\ n^{\varepsilon-1}]} \exp\Big(\frac{-mn}{pq}\delta^2\Big)\mathbb{P}_{\mathcal{B}_{p'}}\big(P\big)d\delta < e^{-n^{11\varepsilon/10}} \int_{[-n^{\varepsilon-1},\ n^{\varepsilon-1}]} \exp\Big(\frac{-mn}{pq}\delta^2\Big)d\delta.$$

We also note that

$$\int_{[-n^{\varepsilon-1},\ n^{\varepsilon-1}]} \exp\Big(\frac{-mn}{pq}\delta^2\Big)d\delta \leq \int_0^1 \exp\Big(\frac{-mn}{pq}\delta^2\Big)d\delta.$$

Hence,

$$\sqrt{\frac{mn}{\pi pq}} \int_{[-n^{\varepsilon-1},\ n^{\varepsilon-1}]} \exp\Big(\frac{-mn}{pq}\delta^2\Big)\mathbb{P}_{\mathcal{B}_{p'}}\big(P\big)d\delta \leq e^{-n^{12\varepsilon/11}}. \tag{7.30}$$

We now bound the value of integral on the intervals $[-p, -\ n^{\varepsilon-1}]$ and $[n^{\varepsilon-1},\ q]$ (which we denote $[-p,\ q] - [-n^{\varepsilon-1},\ n^{\varepsilon-1}]$). Observe that because $\mathcal{B}_p$ is a probability space, $\forall p' \in [0,1],\ \ \mathbb{P}_{\mathcal{B}_{p'}}\big(P\big) \leq 1$ and by definition $\forall p' \notin [0,1],\ \ \mathbb{P}_{\mathcal{B}_{p'}}\big(P\big) = 0$.

Hence by Lemma 9.6 we can bound the 'tails' of the integral as follows,

$$\begin{aligned}\int_{[-p,\,q]-[-n^{\varepsilon-1},\,n^{\varepsilon-1}]} &\exp\Big(\frac{-mn}{pq}\delta^2\Big)\mathbb{P}_{\mathcal{B}_{p'}}\big(P\big)d\delta \\ &= O\bigg(\frac{1}{u}\exp\Big(\frac{-u^2}{2}\Big)\bigg)\int_{[-p,\,q]}\exp\Big(\frac{-mn}{pq}\delta^2\Big)\mathbb{P}_{\mathcal{B}_{p'}}\big(P\big)d\delta \\ &\text{where} \qquad u = \sqrt{\frac{pq}{2mn}}n^{1-\varepsilon} > n^{2\varepsilon/3}.\end{aligned}$$

And thus we have,

$$\sqrt{\frac{mn}{\pi pq}}\int_{[-p,\,p]-[-\varepsilon\ln n,\,\varepsilon\ln n]}\exp\Big(\frac{-mn}{pq}\delta^2\Big)\mathbb{P}_{\mathcal{B}_{p'}}\big(P\big)d\delta \leq e^{-n^{4\varepsilon/3}}. \tag{7.31}$$

The asserted result now follows by (7.30) and (7.31).

□

We are now ready to prove the asymptotic expectation of any random variable in the graph *edge*-model, $\mathcal{G}_M$ can be approximated by its expectation in the binomial *edge*-model, $\mathcal{B}_M$.

Theorem 7.4. *Fix $\varepsilon > 0$ and $0 < a < \frac{1}{2}$ such that $a + \varepsilon < \frac{1}{2}$ and $b + 17\varepsilon < 1$. Let $\mathcal{G}_M = \mathcal{G}_M(m,n)$ be as in Definition 1.7 and let $\mathcal{B}_M = \mathcal{B}_M(m,n)$ be as in Definition 7.2.*

Let $X : I_{m,n,M} \to \mathbb{R}$ be any random variable.

Then as $m,n \to \infty$ subject to Conditions 5.1,

$$\mathbb{E}_{\mathcal{G}_M}\big(X\big) - \mathbb{E}_{\mathcal{G}_M}\big(X\big) = +\,O(n^{-b})\mathbb{E}_{\mathcal{G}_M}\big(|X|\big) + \max_{x\in I_{m,n}}|X(x)|.O\big(e^{-n^{10\varepsilon/11}}\big)$$

Proof. Let $P \subset I_{m,n,M}$ be the set of (a,b,m,n,ε)-*pathological*$(m+n)$-tuples in $I_{m,n,M}$. By the definition of expectation,

$$\mathbb{E}_{\mathcal{G}_M}\big(X\big) = \sum_{x\in P^c}\mathbb{P}_{\mathcal{G}_M}(x)X(x) + \sum_{x\in P}\mathbb{P}_{\mathcal{G}_M}(x)X(x). \tag{7.32}$$

For non-pathological $(\boldsymbol{s},\boldsymbol{t})$, $\mathbb{P}_{\mathcal{G}_M}(\boldsymbol{s},\boldsymbol{t})$ is approximately $\mathbb{P}_{\mathcal{B}_M}(\boldsymbol{s},\boldsymbol{t})$. Hence by Theorem 7.1,

$$\begin{aligned}\mathbb{E}_{\mathcal{G}_M}\big(X\big) &= \sum_{x\in P^c}\mathbb{P}_{\mathcal{B}_M}(x)\big(1+O(n^{-b})\big)X(x) + \max_{x\in I_{m,n}}|X(x)|.O(\mathbb{P}_{\mathcal{G}_M}(P)) &(7.33)\\ &= \sum_{x\in I_{m,n}}\mathbb{P}_{\mathcal{B}_M}(x)\big(1+O(n^{-b})\big)X(x)\\ &\quad + \max_{x\in I_{m,n}}|X(x)|.O\Big(\mathbb{P}_{\mathcal{B}_M}(P)\big(1+O(n^{-b})\big)+\mathbb{P}_{\mathcal{G}_M}(P)\Big). &(7.34)\end{aligned}$$

Consider the first term on the right hand side of (7.34).

$$\begin{aligned}&\sum_{x\in I_{m,n}}\mathbb{P}_{\mathcal{B}_M}(x)\big(1+O(n^{-b})\big)X(x)\\ &\qquad= \sum_{x\in I_{m,n}}\mathbb{P}_{\mathcal{B}_M}(x)X(x) + \sum_{x\in I_{m,n}}\mathbb{P}_{\mathcal{B}_M}(x)O(n^{-b})X(x) &(7.35)\\ &\qquad=\mathbb{E}_{\mathcal{B}_M}\big(X\big) + O(n^{-b})\sum_{x\in I_{m,n}}\mathbb{P}_{\mathcal{B}_M}(x)|X(x)| &(7.36)\\ &\qquad=\mathbb{E}_{\mathcal{B}_M}\big(X\big) + O(n^{-b})\mathbb{E}_{\mathcal{B}_M}\big(|X|\big) &(7.37)\end{aligned}$$

Here, line (7.35) follows by simple rearrangement. Then by the definition of expectation and the uniformity of the error term $O(n^{-b})$ we get (7.36). Then (7.37) follows again by the definition of expectation.

By Theorem 5.2 and Corollary 7.2 we have $\mathbb{P}_{\mathcal{G}_M}(P)\le e^{-n^{11\varepsilon/12}}$ and $\mathbb{P}_{\mathcal{B}_M}(P)\le e^{-n^{12\varepsilon/11}}$ respectively. Hence by (7.34) and (7.37),

$$\mathbb{E}_{\mathcal{G}_M}\big(X\big) =\mathbb{E}_{\mathcal{B}_M}\big(X\big) + O(n^{-b})\mathbb{E}_{\mathcal{B}_M}\big(|X|\big) + \max_{x\in I_{m,n}}|X(x)|.O\big(e^{-n^{10\varepsilon/11}}\big)$$

and we have our result.

□

A similar result holds which relates asymptotic expectations in $\mathcal{G}_p$ and $\mathcal{V}_p$.

Theorem 7.5. *Fix $\varepsilon>0$ and $0<a<\frac{1}{2}$ such that $a+\varepsilon<\frac{1}{2}$ and $b+17\varepsilon<1$. Let $\mathcal{G}_p=\mathcal{G}_p(m,n)$ be as in Definition 1.6 and let $\mathcal{V}_p=\mathcal{V}_p(m,n)$ be as in Definition 7.7.*

Let $X: I_{m,n}\to\mathbb{R}$ be any random variable.

Then as $m,n\to\infty$ subject to Conditions 4.1,

$$\mathbb{E}_{\mathcal{G}_p}\big(X\big)-\mathbb{E}_{\mathcal{V}_p}\big(X\big) = O(n^{-b})\mathbb{E}_{\mathcal{V}_p}\big(|X|\big) + \max_{x\in I_{m,n}}|X(x)|O(e^{-n^{9\varepsilon/10}}).$$

Proof. We begin with the following observations which will be necessary for the proof. Let $y = \frac{\sum_j S_j - pmn}{\sqrt{pqmn}}$ and consider the likely magnitude of $|y|$ in $\mathcal{G}_p$. Note that $y = \sqrt{\frac{mn}{pq}}|\lambda - p|$. Hence by Lemma 4.2, $\mathbb{P}_{\mathcal{G}_p}(|y| > n^{4\varepsilon}) \le e^{-n^{2\varepsilon}}$. (Note that as Lemma 4.2 only concerns the random variables S_j and p (and not T_k's) we also have that $\mathbb{P}_{\mathcal{B}_p}(|y| > n^{4\varepsilon}) \le e^{-n^{2\varepsilon}}$ which implies $\mathbb{P}_{\mathcal{V}_p}(|y| > n^{4\varepsilon}) \le e^{-n^{\varepsilon}}$ in a similar fashion to Corollary 7.3.) The remainder of the proof proceeds in the same way as the proof of Theorem 7.4.

Let $P \subset I_{m,n}$ be the set of $(m+n)$-tuples in $I_{m,n}$ that are either (a,b,m,n,ε)-*pathological* or for which $|y| > 4\varepsilon$. By the definition of expectation,

$$\mathbb{E}_{\mathcal{G}_p}\big(X\big) = \sum_{x \in P^c} \mathbb{P}_{\mathcal{G}_p}(x)X(x) + \sum_{x \in P} \mathbb{P}_{\mathcal{G}_p}(x)X(x). \tag{7.38}$$

By Theorem 7.10, for non-(a,b,m,n,ε)-*pathological* $(\boldsymbol{s},\boldsymbol{t})$, the values $\mathbb{P}_{\mathcal{G}_p}(\boldsymbol{s},\boldsymbol{t})$ and $\mathbb{P}_{\mathcal{V}_p}(\boldsymbol{s},\boldsymbol{t})$ are asymptotically very close. Hence,

$$\mathbb{E}_{\mathcal{G}_p}\big(X\big) = \sum_{x \in P^c} \mathbb{P}_{\mathcal{B}_p}(x)\big(1 + O(n^{-b})\big)X(x) + \max_{x \in I_{m,n}} |X(x)|.O(\mathbb{P}_{\mathcal{G}_p}(P)) \tag{7.39}$$

Similarly to the proof of Theorem 7.4, line (7.39) implies,

$$\begin{aligned} &\mathbb{E}_{\mathcal{G}_p}\big(X\big) - \mathbb{E}_{\mathcal{V}_p}\big(X\big) \\ &\quad = O(n^{-b})\mathbb{E}_{\mathcal{V}_p}\big(|X|\big) + \max_{x \in I_{m,n}} |X(x)|.O\Big(\big(1 + O(n^{-b})\big)\mathbb{P}_{\mathcal{B}_p}\big(P\big) + O(\mathbb{P}_{\mathcal{G}_p}(P))\Big) \end{aligned} \tag{7.40}$$

By Theorem 4.9 and Corollary 7.3 we have $\mathbb{P}_{\mathcal{G}_p}(P) \le e^{-n^{10\varepsilon/11}} + e^{-n^{2\varepsilon}}$ and $\mathbb{P}_{\mathcal{V}_p}(P) \le e^{-n^{\varepsilon}} + e^{-n^{2\varepsilon}}$ respectively. Hence our result follows from (7.40).

□

Consider, any event $A \subset I_{m,n}$ and the random variable I_A defined to be the indicator function of A. Then $\mathbb{E}(I_A) = \mathbb{P}(A)$. Hence Theorems 7.4 and 7.5 yield the following corollaries.

Corollary 7.6. *Under the same conditions and assumptions as Theorem 7.4, let $A \subset I_{m,n,M}$ then,*

$$\mathbb{P}_{\mathcal{G}_M}\big(A\big) - \mathbb{P}_{\mathcal{B}_M}\big(A\big) = O(n^{-b})\mathbb{P}_{\mathcal{B}_M}\big(|A|\big) + O(e^{-n^{10\varepsilon/11}}).$$

Corollary 7.7. *Under the same conditions and assumptions as Theorem 7.5, let $A \subset I_{m,n}$ then,*

$$\mathbb{P}_{\mathcal{G}_p}\big(A\big) - \mathbb{P}_{\mathcal{V}_p}\big(A\big) = O(n^{-b})\mathbb{P}_{\mathcal{V}_p}\big(|A|\big) + O(e^{-n^{9\varepsilon/10}}).$$

We have shown that degree sequences in random bipartite graph models can be approximated by independent binomials subject to certain constraints. The *independence* of the random variables on which our binomial models are based suggests that the degrees of the *white* vertices are perhaps in some sense independent of the degrees of the *black* vertices. We make this idea precise in the following theorem concerning the graph *edge*-model, $\mathcal{G}_M$.

Theorem 7.8. *Fix $a, b \in \mathbb{R}^+$ such that $0 < a + b < \frac{1}{2}$. Then fix $\varepsilon = \varepsilon_0(a, b) > 0$. Let $\mathcal{G}_M = \mathcal{G}_M(m, n)$ be as in Definition 1.7 and $\mathcal{B}_M = \mathcal{B}_M(m, n)$ be as in Definition 7.3.*

For $1 \le j \le m$, let S_j be the random variable that returns the degree *of the* white *vertex u_j and for $1 \le k \le n$, let T_k be the random variable that returns the* degree *of the* black *vertex u_j.*

Let $E_{\boldsymbol{S}}$ be an event defined in terms of the random variables $S_1, \dots, S_m$ and $E_{\boldsymbol{T}}$ be an event defined in terms of the random variables $T_1, \dots, T_n$.

Then as $m, n \to \infty$ subject to Conditions 5.1,

$$\mathbb{P}_{\mathcal{G}_M}\big(E_{\boldsymbol{S}} \text{ and } E_{\boldsymbol{T}}\big) = \big(1 + O(n^{-b})\big)\mathbb{P}_{\mathcal{G}_M}\big(E_{\boldsymbol{S}}\big)\mathbb{P}_{\mathcal{G}_M}\big(E_{\boldsymbol{T}}\big) + O(e^{-n^{10\varepsilon/11}}).$$

Proof.

$$\begin{aligned}\mathbb{P}_{\mathcal{G}_M}&\big(E_{\boldsymbol{S}} \text{ and } E_{\boldsymbol{T}}\big)\\ &= \big(1 + O(n^{-b})\big)\mathbb{P}_{\mathcal{B}_M}\big(E_{\boldsymbol{S}} \text{ and } E_{\boldsymbol{T}}\big) + O(e^{-n^{10\varepsilon/11}}) && (7.41)\\ &= \big(1 + O(n^{-b})\big)\mathbb{P}_{\mathcal{B}_M}(E_{\boldsymbol{S}})\mathbb{P}_{\mathcal{B}_M}(E_{\boldsymbol{T}}) + O(e^{-n^{10\varepsilon/11}}) && (7.42)\\ &= \big(1 + O(n^{-b})\big)\mathbb{P}_{\mathcal{G}_M}(E_{\boldsymbol{S}})\mathbb{P}_{\mathcal{G}_M}(E_{\boldsymbol{T}}) + O(e^{-n^{10\varepsilon/11}}) && (7.43)\end{aligned}$$

We justify the calculation line by line. Define the random variable $I_{E_{\boldsymbol{S}}\&E_{\boldsymbol{T}}}$ to be the indicator function for the event $E_{\boldsymbol{S}}\&E_{\boldsymbol{S}}$. Then the expectation, $\mathbb{E}_{\mathcal{G}_M}(I_{E_{\boldsymbol{S}}\&E_{\boldsymbol{T}}}) = \mathbb{P}_{\mathcal{G}_M}(E_{\boldsymbol{S}}\&E_{\boldsymbol{T}})$. Thus (7.41), follows by Theorem 7.4. Line (7.42) follows by the independence of $\boldsymbol{S}$ and $\boldsymbol{T}$ in the binomial *edge*-model, $\mathcal{B}_M$. Line (7.43) then follows by reapplying Theorem 7.4 to (7.42). And thus we have our result. □

We also observe that such a theorem could not hold in the graph p-model, $\mathcal{G}_p$. Recall that S (resp. T) is the random variable that returns the average degree of the *white* (respectively *black*) vertices.

Now fix an integer M, where $0 \le M \le mn$ and set $E_{\boldsymbol{S}} = \{\boldsymbol{S} \mid \sum_j S_j = M\}$ and $E_{\boldsymbol{T}} = \{T \mid \sum_k T_k = M\}$. Then $E_{\boldsymbol{S}} = \frac{1}{m}M = \frac{n}{m}E_{\boldsymbol{T}}$. Hence for the $\mathcal{G}_p$ model we have a counter-example which shows there is no result in $\mathcal{G}_p$ which is analogous result to Theorem 7.8 in $\mathcal{G}_M$.

Part IV

Appendix

Chapter 8

Calculations on the likelihood of *pathological* degree sequences.

8.1 Generating functions for graph p-model, $\mathcal{G}_p$.

8.1.1 Preliminary expectations

The aim of this section is to provide some calculations needed to show that *pathological* degree sequences are rare in the graph p-model, $\mathcal{G}_p$. We calculate some preliminary expectations on our random variables for the *white* vertices' degrees: $s_1, \ldots, s_m$.

Fix a *white* vertex u_j and consider the degree of that vertex, S_j. By definition of the graph p-model, $\mathcal{G}_p$, the probability that there is an edge between u_j and any *black* vertex v_k is precisely p. As there are n *black* vertices, the random variable S_j is the standard binomial in (n, p). Hence we can write probability generating function, $A_1(x)$, for S_j in $\mathcal{G}_p$.

$$A_1(x) = (px + q)^n$$

This generating function will allow us to calculate the expectation of S_j and of S_j^2. By the theory of generating functions explained in Section 3.3 this requires the evaluation of some differential operators on the function $A_1(x)$.

8.1.2 Differentials of A_1, $(\mathcal{G}_p)$.

$$
\begin{aligned}
(xD)A_1 &= xpn(px+q)^{n-1} \\
\big((xD)A_1\big)_{x=1} &= pn \qquad (8.1)
\end{aligned}
$$

$$
\begin{aligned}
(xD)^2A_1 &= x^2D^2A_1 + xDA_1 = x^2p^2n(n-1)(px+q)^{n-2} + xpn(px+q)^{n-1} \\
\big((xD)^2A_1\big)_{x=1} &= p^2n(n-1) + pn = pn(1-p) + p^2n^2 = pnq + p^2n^2 \qquad (8.2)
\end{aligned}
$$

We can now calculate the following expectations of the functions on the degrees of the vertices. Note these hold for all $j = 1, \ldots, m$. The first two lines, the results on A_1, are from the lines (8.1) and (8.2).

$$
\begin{aligned}
\mathbb{E}[S_j] &= \big(\frac{d}{dx}A_1(x)\big)_{x=1} = \big(np(px+q)^{n-1}\big)_{x=1} = np \\
\mathbb{E}[S_j^2] &= \big(\frac{d}{dx}x\frac{d}{dx}A_1(x)\big)_{x=1} = \big(\frac{d}{dx}xnp(px+q)^{n-1}\big)_{x=1} = np + n(n-1)p^2
\end{aligned}
$$

Thus by linearity of expectation,

$$
\begin{aligned}
\mathbb{E}[\sum_j S_j^2] &= \sum_j \mathbb{E}[S_j^2] = mnp + mn(n-1)p^2 \\
\mathbb{E}[\sum_j S_j] &= \sum_j \mathbb{E}[S_j] = mnp
\end{aligned}
$$

8.1.3 Expectation of $\sum_j (S_j - S)^2$

$$
\begin{aligned}
\mathbb{E}[\sum_j (S_j - S)^2] &= \mathbb{E}[\sum_j S_j^2] - 2\mathbb{E}[S\sum_j S_j] + m\mathbb{E}[S^2] \\
&= \mathbb{E}[\sum_j S_j^2] - \frac{2}{m}\mathbb{E}[(\sum_j S_j)^2] + \frac{m}{m^2}\mathbb{E}[(\sum_j S_j)^2] \\
&= \mathbb{E}[\sum_j S_j^2] - \frac{1}{m}\mathbb{E}[(\sum_j S_j)^2] \\
&= mnp + mn(n-1)p^2 - \frac{1}{m}(mn(mn-1)p^2 + mnp) \\
&= mnpq - npq \qquad (8.3)
\end{aligned}
$$

8.1.4 Expectation of $\sum_j (S_j - np)^2$

$$
\begin{aligned}
\mathbb{E}[\sum_j (S_j - np)^2] &= \mathbb{E}[S_j^2] - 2\mathbb{E}[np \sum_j S_j] + mn^2p^2 \\
&= mnp + mn(n-1)p^2 - 2np(mnp) + mn^2p^2 \\
&= mnpq
\end{aligned} \tag{8.4}
$$

8.2 Generating functions for graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$.

8.2.1 Differentials of B, ($\mathcal{G}_{\boldsymbol{t}}$)

We work in the random graph model $\mathcal{G}_{\boldsymbol{t}}$. In this model the degrees of the *black* vertices, i.e. $\boldsymbol{t}$, are given, and all bipartite graphs with this *black* degree sequence are equally likely. This section provides some calculations needed to show that *pathological* degree sequences are rare in the graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$.

Recall that $B(x)$ is the probability generating function for S_j, the degree of vertex u_j in random graph model $\mathcal{G}_{\boldsymbol{t}}$. Similar to the previous graph model $\mathcal{G}_p$, we find the expectation of $\sum_j (S_j - S)^2$ in $\mathcal{G}_{\boldsymbol{t}}$. For this purpose we calculate up to the second order derivative of $B(x)$.

$$
\begin{aligned}
B &= \prod_{k=1}^{n} (\frac{t_k}{m}x + \frac{m - t_k}{m}) \\
DB &= \frac{1}{m}(\sum_{h=1}^{n} t_h \prod_{k \neq h} (\frac{t_k}{m}x + \frac{m - t_k}{m})) \\
\big(\, DB(x) \,\big)_{x=1} &= m^{-1} \sum_k t_k = \lambda n && (8.5) \\
D^2B(x) &= \frac{1}{m^2}(\sum_{h \neq l} t_h t_l \prod_{k \notin \{h,l\}} (\frac{t_k}{m}x + \frac{m - t_k}{m})) \\
\big(\, D^2B(x) \,\big)_{x=1} &= m^{-2} \sum_{h \neq l} t_h t_l = m^{-2}\big((\sum_k t_k)^2 - \sum_k t_k^2\big) = \lambda^2 n^2 - m^{-2} \sum_k t_k^2 \\
\big((xD)^2 B\big)_{x=1} &= \big(x^2 D^2 B + xDB\big)_{x=1} = \lambda n + \lambda^2 n^2 - m^{-2} \sum_k t_k^2 && (8.6)
\end{aligned}
$$

8.3 Differential operators and symbolic D notation

This section is not strictly necessary for the thesis but is included here for the interested reader. Let $A(x)$ be the probability generating function for some random variable X. Then the calculation of $\big((xD)^n A(x)\big)_{x=1}$ for large $n \in N$, i.e. $\mathbb{E}(X^n)$ can be streamlined by noting the following relation.

$$Dx^iD = ix^{i-1}D + x^iD^2$$

$$\begin{aligned}
(xD)^2 &= xDxD = x(xD^2 + D) = x^2D^2 + xD \\
(xD)^3 &= xD(xD + x^2D) = xD + 3x^2D^2 + x^3D^3 \\
(xD)^4 &= xD(xD + 3x^2D^2 + x^3D^3) = xD + 7x^2D^2 + 6x^3D^3 + x^4D^4 \\
(xD)^5 &= xD(xD + 7x^2D^2 + 6x^3D^3 + x^4D^4) = xD + 15x^2D^2 + 25x^3D^3 + 10x^4D^4 + x^5D^5
\end{aligned}$$

The coefficients above form the following pattern.

1						
1	1					
1	3	1				
1	7	6	1			
1	15	25	10	1		
1	31	90	65	15	1	
1	63	301	350	140	21	1

These are the Stirling numbers of the second kind and are defined recursively by,

$$\binom{a}{a} = \binom{a}{0} = 1, \qquad \binom{i}{j} = \binom{i-1}{j-1} + (j+1)\binom{i-1}{j-1}.$$

Chapter 9

Results needed for Chapter 7.

Lemmas 7.8 and 7.9 form part of our proof that non-*pathological* degree sequences have similar probabilities under the $\mathcal{G}_p$ and $\mathcal{V}_p$ models. We prove these results first.

The other result we will need to prove is Lemma 7.3 which is used to calculate the asymptotic probability of choosing a particular $(m+n)$-tuple $(\boldsymbol{s}, \boldsymbol{t})$ in the binomial p-model, $\mathcal{B}_p$. We prove this result in Section 9.2.

9.1 Proof of Lemmas 7.8 and 7.9.

Asymptotic combinatorial results Some preliminary results from asymptotic combinatorics are needed to show Lemmas 7.8 and 7.9. We state and sometimes re-develop these results below for the interested reader. The first results are stated without proof. Stirling's approximation is often used to bound factorials. A reference can be found on line (4.2) of [Odl95, p.1076].

Lemma 9.1 (Stirling's approximation)**.**

$$n! = \sqrt{2\pi n}\Big(\frac{n}{e}\Big)^n\bigg(1 + O\Big(\frac{1}{n}\Big)\bigg).$$

The next lemma concerns rising and falling factorial approximations. This is an often used result but we give a proof here for completeness.

Lemma 9.2. *Let $k \in \mathbb{R}$ and assume $|\frac{k}{n}| < c$ for some $c < 1$. Then,*
Note that we can now write for any $k \in \mathbb{R}$:

1.
$$\frac{(n+k)!}{n!\,n^k} = \exp\big(\frac{k^2}{2n} + O(\frac{k}{n} + \frac{k^3}{n^2})\big)$$

2.
$$\frac{(n-k)!}{n!\,n^{-k}} = \exp\big(\frac{k^2}{2n} + O(\frac{k}{n} + \frac{k^3}{n^2})\big)$$

Proof. (of 1.) We use the trick of writing $x = e^{\ln x}$ and then applying the series expansion for natural log.

$$\begin{aligned}\frac{(n+k)!}{n!\,n^k} &= \prod_{i=1}^{k}\Big(1 + \frac{i}{n}\Big)\\ &= \exp\Bigg(\sum_{i=1}^{k} \ln\Big(1+\frac{i}{n}\Big)\Bigg)\\ &= \exp\Bigg(\sum_{i=1}^{k} \frac{i}{n} + O\Big(\frac{i^2}{n^2}\Big)\Bigg)\\ &= \exp\Bigg(\frac{k^2}{2n} + O\Big(\frac{k}{n} + \frac{k^3}{n^2}\Big)\Bigg).\end{aligned}$$
□

Proof. (of 2.) This proceeds similarly to the proof of part (1). □

Proof of Lemma 7.8. We will prove Lemma 7.8 first and proceed by a series of preliminary lemmas.

Lemma 9.3. *Fix $0 < a < \frac{1}{2}$. Suppose p is* acceptable *for a, m and n and write $q = 1 - p$. Then,*
$$(p^{pn}q^{qn})^{-1}\frac{(pn)!(qn)!}{n!} = \sqrt{2\pi pqn}\Bigg(1 + O\Big(\frac{\ln n}{n}\Big)\Bigg)$$

Proof. In the calculations below, (9.1) follows by Stirling's approximation (Lemma 9.1). Many terms in (9.1) then cancel to give us (9.2) as required.

$$\begin{aligned}&(p^{pn}q^{qn})^{-1}\frac{(pn)!(qn)!}{n!}\\ &= (p^{pn}q^{qn})^{-1}\sqrt{2\pi}\sqrt{\frac{pnqn}{n}}\exp\big(n - pn - qn\big)\frac{(pn)^{pn}(qn)^{qn}}{n^n}\Bigg(1 + O\Big(\frac{1}{pn} + \frac{1}{qn}\Big)\Bigg) && (9.1)\\ &= \sqrt{2\pi pqn}\Bigg(1 + O\Big(\frac{\ln n}{n}\Big)\Bigg). && (9.2)\end{aligned}$$

□

Lemma 9.4. *Define* $y := \frac{M-pmn}{\sqrt{pqmn}}$ *and assume that* $|y| < n^{\varepsilon}$. *Fix* $0 < a < \frac{1}{2}$. *Write* $q = 1-p$. *Then as* $m, n \to \infty$ *subject to* p acceptable *for* a, m *and* n,

$$\frac{(pmn + y\sqrt{pqmn})!(qmn - y\sqrt{pqmn})!}{(pmn)!(qmn)!p^{y\sqrt{pqmn}}q^{-y\sqrt{pqmn}}} = \exp\left(\frac{y^2}{2} + O\big((mn)^{-1/2+4\varepsilon}\big)\right)$$

Proof. Note it is sufficient to prove the following two approximations.

$$\frac{(pmn + y\sqrt{pqmn})!}{(pmn)!(pmn)^{y\sqrt{pqmn}}} = \exp\left(\frac{qy^2}{2} + O\big((mn)^{-1/2+4\varepsilon}\big)\right) \tag{9.3}$$

$$\frac{(qmn - y\sqrt{pqmn})!}{(qmn)!(qmn)^{-y\sqrt{pqmn}}} = \exp\left(\frac{py^2}{2} + O\big((mn)^{-1/2+4\varepsilon}\big)\right). \tag{9.4}$$

In the proofs of both of these results we use the approximations developed for rising and falling factorials in Lemma 9.2. The last line in each of the calculations below follows by our assumption that $|y| < n^{\varepsilon}$.

We begin with the proof of (9.3).

$$\begin{aligned}\frac{(pmn + y\sqrt{pqmn})!}{(pmn)!(pmn)^{y\sqrt{pqmn}}} &= \exp\big(\frac{y^2pqmn}{2pmn} + O\big(\frac{y\sqrt{pqmn}}{pmn} + \frac{y^3(pqmn)^{3/2}}{(pmn)^2}\big)\big)\\ &= \exp\big(\frac{y^2q}{2} + O\big(\frac{y\sqrt{q}}{\sqrt{pmn}} + \frac{y^3q^{3/2}}{\sqrt{pmn}}\big)\big)\\ &= \exp\big(\frac{y^2q}{2} + O\big((mn)^{-1/2+4\varepsilon}\big)\big)\end{aligned}$$

The proof of (9.4) proceeds similarly. □

We are now in a position to prove Lemma 7.8. This lemma was originally stated on p84 of this thesis but we restate it below for convenience.

Lemma 7.8. *We define* $y := \frac{M-pmn}{\sqrt{pqmn}}$ *and assume that* $|y| < n^{\varepsilon}$. Then as $m, n \to \infty$ subject to p *acceptable* for a, m and n,

$$\left(p^M q^{mn-M}\binom{mn}{M}\right)^{-1} = \sqrt{2\pi pqmn}\exp\left(\frac{y^2}{2} + O\big((mn)^{-1/2+\varepsilon}\big)\right).$$

Proof. Observe that by the definition of y,

$$M = y\sqrt{pqmn} + pmn \qquad \text{and} \qquad M - mn = y\sqrt{pqmn} - qmn.$$

Hence by expansion and some rearrangement,

$$\left(p^M q^{mn-M}\binom{mn}{M}\right)^{-1} = \frac{(pmn)!(qmn)!}{(mn)!p^{pmn}q^{qmn}} \times \frac{(pmn + y\sqrt{pqmn})!(qmn - y\sqrt{pqmn})!}{(pmn)!(qmn)!p^{y\sqrt{pqmn}}q^{-y\sqrt{pqmn}}}$$

In the expression above we can approximate the two fractions on the right hand side by Lemmas 9.3 and 9.4 respectively. This yields,

$$\big(p^M q^{mn-M}\binom{mn}{M}\big)^{-1} = \sqrt{2\pi pqmn}\big(1 + O\big(\tfrac{1}{mn}\big)\big) \exp\left(\frac{y^2}{2} + O\Big((mn)^{-1/2+\varepsilon}\Big)\right)$$

as required. $\square$

Proof of Lemma 7.9. In the first step in the proof of this result, we find an exponential approximation to $\big(\frac{p'}{p}\big)^{2M+\frac{1}{2}}\big(\frac{q'}{q}\big)^{2mn-2M+\frac{1}{2}}$.

Lemma 9.5. *Fix $\varepsilon > 0$ and set $\delta = p - p'$. Also assume $|\delta| < n^{\varepsilon}\sqrt{\frac{pq}{mn}}$. Write $q = 1 - p$ and $q' = 1 - p'$. Then as $m, n \to \infty$ subject to Conditions 4.1,*

$$\begin{aligned}
&\left(\frac{p'}{p}\right)^{2M+\frac{1}{2}}\left(\frac{q'}{q}\right)^{2mn-2M+\frac{1}{2}} \\
&\qquad = \exp\Big(\big(2pmn + 2y\sqrt{pqmn} + \tfrac{1}{2}\big)\big(\tfrac{1}{p}\delta - \tfrac{1}{2p^2}\delta^2\big)\Big) \\
&\qquad\quad \times \exp\Big(\big(2qmn - 2y\sqrt{pqmn} + \tfrac{1}{2}\big)\big(-\tfrac{1}{q}\delta - \tfrac{1}{2q^2}\delta^2\big)\Big)\big(1 + O(n^{-1+10\varepsilon})\big).
\end{aligned}$$

Proof. Similar to the proof of 9.2 we will use the trick of writing $x = e^{\ln x}$ and then applying the series expansion for natural log. This method is first applied to $\frac{p'}{p}$,

$$\begin{aligned}\frac{p'}{p} =& 1+\frac{\delta}{p} = \exp\Big(\ln\big(1+\frac{\delta}{p}\big)\Big)\\ =& \exp\Big(\frac{\delta}{p}-\frac{\delta^2}{2p^2}+O\big(\frac{\delta^3}{p^3}\big)\Big)\end{aligned}$$

Similarly,

$$\begin{aligned}\frac{q'}{q} =& 1-\frac{\delta}{q} = \exp\Big(\ln\big(1+\frac{\delta}{q}\big)\Big)\\ =& \exp\Big(-\frac{\delta}{q}-\frac{\delta^2}{2q^2}+O\big(\frac{\delta^3}{p^3}\big)\Big).\end{aligned}$$

Hence we have,

$$\begin{aligned}&\Big(\frac{p'}{p}\Big)^{2M+\frac{1}{2}}\Big(\frac{q'}{q}\Big)^{2mn-2M+\frac{1}{2}}\\ &\quad=\exp\Big((2pmn+2y\sqrt{pqmn}+\tfrac{1}{2})\big(\tfrac{1}{p}\delta-\tfrac{1}{2p^2}\delta^2\big)\Big)\\ &\quad\quad\times\exp\Big((2qmn-2y\sqrt{pqmn}+\tfrac{1}{2})\big(-\tfrac{1}{q}\delta-\tfrac{1}{2q^2}\delta^2\big)\Big)\\ &\quad\quad\times\exp\Big(O\big((2qmn-2y\sqrt{pqmn}+\tfrac{1}{2})\tfrac{\delta^3}{p^3}\big)\Big).\end{aligned}\tag{9.5}$$

To bound the error term first note that by our assumptions on $|\delta|$ we have,

$$\frac{\delta^3}{p^3}<\frac{n^{3\varepsilon}}{p^3}\Big(\sqrt{\frac{pq}{mn}}\Big)^3=n^{3\varepsilon}(mn)^{-3/2}\sqrt{\frac{q^3}{p^3}}.\tag{9.6}$$

The restrictions on m,n imply that $(mn)^{-3/2}<n^{-3+4\varepsilon}$. Also by Lemma 4.5, $p>\frac{1}{\ln n}$ and so in particular, $\sqrt{\frac{q^3}{p^3}}<(\ln n)^{3/2}$. Hence by (9.6),

$$\frac{\delta^3}{p^3}<n^{3\varepsilon}n^{-3+4\varepsilon}(\ln n)^{3/2}<n^{-3+8\varepsilon}.\tag{9.7}$$

The result now follows by (9.5) and (9.7). □

This Lemma appears, for example, in the preliminary theory of Bollobás' book [Bol01].

Lemma 9.6.

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-u\sigma}^{u\sigma}\exp\big(-x^2/2\sigma^2\big)dx=1+O\Big(\frac{1}{u}\exp\big(-u^2/2\big)\Big)$$

In the following Corollary we put Lemma 9.6 into a more convenient form for our calculations.

Corollary 9.7. *Let* $u \leq \frac{1}{\sqrt{2a}} \times \min\{k - \frac{b}{2a}, k + \frac{b}{2a}\}$. *Then,*

$$\int_{-k}^{k} \exp\big(-a\delta^2 + b\delta\big)dx = \sqrt{\frac{\pi}{a}} \exp\Big(\frac{b^2}{4a}\Big)\Big(1 + O\Big(\frac{1}{u}\exp\big(-u^2/2\big)\Big)\Big)$$

We are now ready to prove Lemma 7.9. It is restated here for convenience.

Lemma 7.9.
Fix $\varepsilon > 0$. Let $y = \frac{M - pmn}{\sqrt{pqmn}}$ and $\delta = p - p'$. Write $q = 1 - p$ and $q' = 1 - p'$. Assume $|y| < n^{4\varepsilon}$. Then as $m, n \to \infty$ subject to p *acceptable* for a, m and n,

$$\int_{-\infty}^{\infty} \big(\frac{p'}{p}\big)^{2M+1/2}\big(\frac{q'}{q}\big)^{2mn-2M+1/2} \exp\Big(-\frac{mn}{pq}\delta^2\Big)d\delta = \sqrt{\frac{\pi pq}{2mn}} \exp\Big(\frac{y^2}{2}\Big)\Big(1 + O(n^{-1/2+12\varepsilon})\Big)$$

Proof. The calculations begin, we work to simplify the integral.

$$\int_{-n^{5\varepsilon}}^{n^{5\varepsilon}} \big(\frac{p'}{p}\big)^{2M+1/2}\big(\frac{1-p'}{1-p}\big)^{2mn-2M+1/2} \exp\big(\frac{-mn}{pq}(p-p')^2\big) \ d\delta \tag{9.8}$$

$$\begin{aligned} &= \int_{\delta=-\ln n}^{\delta=\ln n} \exp\big(\frac{-mn}{pq}\delta^2\big) \exp\big((2pmn + 2y\sqrt{pqmn} + \tfrac{1}{2})\big(\tfrac{1}{p}\delta - \tfrac{1}{2p^2}\delta^2\big)\big) \\ &\quad \exp\big((2qmn - 2y\sqrt{pqmn} + \tfrac{1}{2})\big(-\tfrac{1}{q}\delta - \tfrac{1}{2q^2}\delta^2\big)\big)\big(1 + O(-n^{-1+10\varepsilon})\big) \end{aligned} \tag{9.9}$$

$$\begin{aligned} &= (1 + O(-n^{-1+10\varepsilon})) \int_{\delta=-\ln n}^{\delta=\ln n} \exp\big(-a\delta^2 + b\delta\big) \\ &\quad \text{where,} \\ &\quad a = \tfrac{1}{(2pq)^2}(8mnpq + 4y\sqrt{pqmn}(1-2p) + 2p^2 - 2p + 1), \\ &\quad b = \tfrac{1}{2pq}(4y\sqrt{pqmn} - (2p-1)). \end{aligned} \tag{9.10}$$

Line 9.9 follows by Lemma 9.5. This then rearranges to give line 9.10. We now find an approximation for $\frac{b^2}{4a}$. By (9.10),

$$\begin{aligned}
\frac{b^2}{4a} &= \frac{(4y\sqrt{pqmn}-(2p-1))^2}{4(8mnpq+4y\sqrt{pqmn}(1-2p)+2p^2-2p+1)} \\
&= \frac{16y^2pqmn-(2p-1)8y\sqrt{pqmn}+(2p-1)^2}{4(8mnpq+4y\sqrt{pqmn}(1-2p)+2p^2-2p+1)} \\
&= \frac{y^2}{2} + \frac{((1-2p)8y\sqrt{pqmn}(y-1)+(2p-1)^2-y(2p^2-2p+1)}{4(8mnpq+4y\sqrt{pqmn}(1-2p)+2p^2-2p+1)}
\end{aligned}$$

thus,

$$\frac{b^2}{4a} = \frac{y^2}{2} + O((mn)^{-1/2+2\varepsilon}). \tag{9.11}$$

We note also that by (9.10), $a = \frac{2mn}{pq} + O((mn)^{-1/2+2\varepsilon})$. Then by Corollary 9.7 we have,

$$\begin{aligned}
&\int_{-n^{5\varepsilon}}^{n^{5\varepsilon}} \big(\frac{p'}{p}\big)^{2M+1/2}\big(\frac{1-p'}{1-p}\big)^{2mn-2M+1/2} \exp\big(\frac{-mn}{pq}(p-p')^2\big)\ d\delta \\
&= \sqrt{\frac{\pi pq}{2mn} + O((mn)^{-1/2+2\varepsilon})} \exp\Big(\frac{y^2}{2} + O((mn)^{-1/2+2\varepsilon})\Big) \\
&\quad \times \Big(1 + O\Big(\frac{1}{u}\exp\big(-u^2/2\big)\Big)\Big) \\
&= \Big(\sqrt{\frac{\pi pq}{2mn}}\exp\Big(\frac{y^2}{2}\Big)(1+O(n^{-1+10\varepsilon})) + O(n^{-1/2+4\varepsilon})\Big) \\
&\quad \times \Big(1 + O\Big(\frac{1}{u}\exp\big(-u^2/2\big)\Big)\Big) \\
&\quad \text{where,} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (9.12) \\
&\quad u = \sqrt{\frac{pq}{4mn}} \times \Big(n^{5\varepsilon} - y + O(n^{-1/2+4\varepsilon})\Big) > n^{4\varepsilon}
\end{aligned}$$

The right hand side of (9.12) then simplifies to yield,

$$\sqrt{\frac{\pi pq}{2mn}}\exp\Big(\frac{y^2}{2}\Big) + O\big(n^{-1+12\varepsilon}\big)$$

and so we are done.

□

9.2 Proof of Lemma 7.3

We state the Lemma again for convenience and provide the proof.

Lemma 7.3 Let $0 < a < 1/2$ and suppose that $pq > \frac{1}{\ln n}$. Then,

$$\sum_{j=0}^{n}\left(\binom{n}{j}p^j q^{n-j}\right)^2 = \frac{1}{2\sqrt{\pi npq}}\Big(1 + O\left(n^{-1/2+\varepsilon}\right)\Big) + O(e^{-n^{\varepsilon}})$$

Proof. Consider the region where the greater proportion of the sum lies. The largest term in the sum lies around $j = np$. Make the substitution $j = np + x$. and split the sum into two components: $|j - np| < n^{3/5}$ and $|j - np| \leq n^{3/5}$.

$$\begin{aligned}&\sum_{j=0}^{n}\left(\binom{n}{j}p^j q^{n-j}\right)^2\\ &\quad = \sum_{|np-j|<n^{3/5}}\left(\binom{n}{j}p^j q^{n-j}\right)^2 + \sum_{|np-j|\geq n^{3/5}}\left(\binom{n}{j}p^j q^{n-j}\right)^2\end{aligned}$$

We first write a general outline of how the proof will proceed, writing $\boldsymbol{E_1}, \ldots$ for any error terms we pick up. We will then bound each of these errors in turn. Thus,

$$\sum_{j=0}^{n}\left(\binom{n}{j}p^j q^{n-j}\right)^2 = \sum_{|np-j|<n^{3/5}}\left(\binom{n}{j}p^j q^{n-j}\right)^2 + \boldsymbol{E_1}$$

where $\boldsymbol{E_1}$ denotes the 'tails' of the sum. We can now use the Euler-Maclaurin summation formula to approximate the central part of the sum with a truncated normal distribution. Make the substitution $x = np - j$ and write,

$$\int_{j=0}^{n}\left(\binom{n}{j}p^j q^{n-j}\right)^2 = \sum_{|x|<n^{3/5}}\frac{1}{2\pi npq}\exp\Big(-\frac{x^2}{pqn}\Big)(1 + \boldsymbol{E_2}) + \boldsymbol{E_1}.$$

Now, because the tails of the normal distribution are small,

$$\sum_{j=0}^{n}\left(\binom{n}{j}p^j q^{n-j}\right)^2 = \boldsymbol{E_3} + \int_{-\infty}^{\infty}\frac{1}{2\pi npq}\exp(-\frac{x^2}{pqn})(1 + \boldsymbol{E_2}) + \boldsymbol{E_1}.$$

After integrating,

$$\sum_{j=0}^{n}\left(\binom{n}{j}p^j q^{n-j}\right)^2 = \boldsymbol{E_3} + \frac{1}{2\sqrt{\pi npq}}(1 + \boldsymbol{E_2}) + \boldsymbol{E_1}.$$

It now remains to bound $\boldsymbol{E_1}, \boldsymbol{E_2}, \boldsymbol{E_3}$ and $\boldsymbol{E_4}$. We begin by bounding $\boldsymbol{E_2}$. We calculate, using Stirling's approximation for factorials (Lemma 9.1), that,

$$\begin{aligned}
&\left(\binom{n}{np+x}p^{np+x}q^{nq-x}\right)^2 \\
&= \frac{n}{2\pi(np+x)(nq-x)}\left(\frac{np}{np+x}\right)^{2(np+x)}\left(\frac{nq}{nq-x}\right)^{2(nq-x)}\left(1+O\left(\frac{\ln n}{n}\right)\right)
\end{aligned}$$

Where the error term follows because p is acc and thus $p > \frac{1}{\ln n}$. Then use the trick of writing $z = e^{\ln z}$ and using the series expansion for natural log. This yields

$$\begin{aligned}
&\left(\binom{n}{np+x}p^{np+x}q^{nq-x}\right)^2 \\
&= \frac{1}{2\pi npq}\exp\left(-2(np+x)\log(1+\frac{x}{np}) - 2(nq-x)\log(1-\frac{x}{nq})\right)\left(1+O\left(\frac{x}{n}\right)\right)
\end{aligned}$$

which simplifies to,

$$\left(\binom{n}{np+x}p^{np+x}q^{nq-x}\right)^2 = \frac{1}{2\pi npq}\exp\left(\frac{-x^2}{npq}\right)\left(1+O\left(n^{-1/5}\right)\right). \tag{9.13}$$

$$\tag{9.14}$$

Where the bound above is uniform. Thus the term,

$$\sum_{|np-j|<n^{3/5}}\left(\binom{n}{j}p^j q^{n-j}\right)^2$$

is the sum of a Gaussian. We now approximate this sum with an integral. To do this we apply the Euler-Maclaurin summation (for the definition of the terms B_2, R_2 etc. see [Odl95, p.1090]). The Euler-Maclaurin result means the summation over a function g can be approximated by an integral as follows,

$$\sum_{k=a}^{b} g(k) = \int_a^b g(x)dx + \frac{B_2}{2}\left(g'(b) - g'(a)\right) + \frac{1}{2}(g(a)+g(b)) + R_2.$$

In our case the error term will be R_2, so we calculate the magnitude of this,

$$
\begin{aligned}
|R_2| \leq& \frac{|B_4|}{4!}\int_{-(pqn)^{1/2+\varepsilon}}^{(pqn)^{1/2+\varepsilon}} |g''''(x)|dx \\
\leq& \frac{1}{6!}\int_{-\infty}^{\infty} |g''''(x)|dx \\
=& \frac{1}{6!}\sqrt{\pi npq}\left(\frac{12}{(pqn)^2} + 2npq\frac{48}{(pqn)^3} + \tfrac{3.16}{4}\frac{1}{(pqn)^4}(npq)^2\right) \\
=& O(n^{-1/2})
\end{aligned}
$$

Hence we now have,

$$
\sum_{|x|<n^{\frac{3}{5}}} \left(\tbinom{n}{np+x}p^{np+x}q^{nq-x}\right)^2 = \frac{1}{2\pi npq}\left(1+O\left(n^{-1/2+\varepsilon}\right)\right)\int_{|x|<n^{\frac{3}{5}}} \exp\left(\frac{-x^2}{npq}\right)dx + O(e^{-n^{\varepsilon}})
$$

And so $\boldsymbol{E_2} = O(n^{-1/2+\varepsilon})$. One of our intermediate results to bound $\boldsymbol{E_2}$ can be used to bound $\boldsymbol{E_1}$.

The error $\boldsymbol{E_1}$ corresponds to the tails of a binomial sum. As we know that the binomal distribution always decreases away from its mean, then we can bound $\boldsymbol{E_1}$ by n times its maximum height.

$$
\begin{aligned}
\boldsymbol{E_1} =& \left|\sum_{|x|\geq n^{\frac{3}{5}}} \left(\binom{n}{j}p^j(1-p)^{n-j}\right)^2\right| \\
\leq& n\left|\frac{1}{2\pi npq}\exp\left(\frac{-n^{3/52}}{npq}\right)\right| \\
=& \frac{1}{2\pi pq}\exp\left(-n^{6/5+\varepsilon}\right)
\end{aligned}
$$

The error $\boldsymbol{E_3}$ corresponds to the tails of normal distribution. Thus we can bound it by Lemma 9.6. The magnitude of its error is absorbed into the other error terms. And so we are done. □

Glossary of Notations

Graph parameters See page v of the introduction for complete definitions.

m	$\#$*white* vertices.
n	$\#$*black* vertices.
u_j	j^{th} *white* vertex.
v_k	k^{th} *black* vertex.
s_j	degree of u_j.
$\boldsymbol{s}$	$s_1, \ldots, s_m$.
s	$\frac{1}{m}\sum_j s_j$.
t_k	degree of v_k.
$\boldsymbol{t}$	$t_1, \ldots, t_k$.
t	$\frac{1}{n}\sum_k t_k$.
λ	edge density, defined to be $\lambda = \frac{1}{mn}\sum_j s_j$.

Random variables on bipartite graphs In each of our random graph models we consider the properties of random variables defined on that space. Generally uppercase letters are used for the random variables corresponding to the deterministic parameters of bipartite graphs which are denoted by their lowercase counterparts.

S_j	degree of u_j.
S_j^*	truncated degree, see Definition 4.1.
$\boldsymbol{S}$	$S_1, \ldots, S_m$.
S	$\frac{1}{m}\sum_j s_j$.
T_k	degree of v_k.
$\boldsymbol{T}$	$T_1, \ldots, T_k$.
T	$\frac{1}{n}\sum_k t_k$.
λ	edge density, defined to be $\lambda = \frac{1}{mn}\sum_j S_j$.

Notation and terms defined to bound the likelihood of *pathological* degree sequences $\mathcal{G}_{\boldsymbol{t}}$.

locally ordered	see Definition 6.2.
$\mathcal{A}, \mathcal{A}_l$	see Definition 6.3.
reference function $\mathcal{R}$	see Definition 6.4.
$Y_l^{\boldsymbol{a}}$	see Definition 6.5.
toxic (node)	see Definition 6.11.
bad (node)	see Definition 6.13.
good (node)	if not *bad*.
martingale X_i	$X_i := \mathbb{E}_{\mathcal{G}_{\boldsymbol{t}}^{\varnothing}}\big(\sum_j (S_j - \lambda n)^2 \mid \mathcal{F}_i\big)$, see Definition 4.1.
σ-algebra $\mathcal{F}_l$	is the σ-algebra induced by the partition $\{Y_l^{\boldsymbol{a}}\}_{\boldsymbol{a} \in \mathcal{A}_l}$, see Definition 6.4

Probability Spaces on Bipartite Graphs

Graph p-model, $\mathcal{G}_p$.	Definition 1.6 on p5
Graph *edge*-model, $\mathcal{G}_M$.	Definition 1.7 on p5
Graph *half*-model, $\mathcal{G}_{\boldsymbol{t}}$.	Definition 1.8 on p6
Graph *ordered-half*-model, $\mathcal{G}_{\boldsymbol{t}}^{\boldsymbol{a}}$.	Definition 6.7 on p57

Binomially based Probability Spaces

Binomial *integrated*-model, $\mathcal{I}_p$.	Definition 7.1 on p73
Binomial p-model, $\mathcal{B}_p$.	Definition 7.2 on p74
Binomial *edge*-model, $\mathcal{B}_M$.	Definition 7.3 on p74
Binomial *half*-model, $\mathcal{B}_{\boldsymbol{t}}$.	Definition 7.5 on p76
Binomial *integrated*-model, $\mathcal{V}_p$.	Definition 7.7 on p84

Terminology

acceptable for a, m and n	Definition 2.5 on p23
(ε, a)-*regular*	Definition 2.7 on p23
$(a, b, m, n, \varepsilon)$-*pathological*	Definition 2.9 on p24

Bibliography

[Azu67] Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tôhoku Math. J. (2)*, 19:357–367, 1967.

[Bar10] Alexander Barvinok. On the number of matrices and a random matrix with prescribed row and column sums and 0–1 entries. *Adv. Math.*, 224(1):316–339, 2010.

[Bol82] Béla Bollobás. Vertices of given degree in a random graph. *J. Graph Theory*, 6(2):147–155, 1982.

[Bol98] Béla Bollobás. *Modern graph theory*, volume 184 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1998.

[Bol01] Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[BT87] B. Bollobás and A. Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.

[Che81] Herman Chernoff. A note on an inequality involving the normal distribution. *Ann. Probab.*, 9(3):533–535, 1981.

[CL06] Fan Chung and Linyuan Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Math.*, 3(1):79–127, 2006.

[CMZ09] Colin Cooper, Andrew McGrae, and Michele Zito. Martingales on trees and the empire chromatic number of random trees. In Miroslaw Kutylowski, Witold Charatonik, and Maciej Gebala, editors, *Fundamentals of Computation Theory*, volume 5699 of *Lecture Notes in Computer Science*, pages 74–83. Springer Berlin / Heidelberg, 2009.

[EKR76] P. Erdős, D. J. Kleitman, and B. L. Rothschild. Asymptotic enumeration of K_n-free graphs. In *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II*, pages 19–27. Atti dei Convegni Lincei, No. 17. Accad. Naz. Lincei, Rome, 1976.

[ER59] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[ER61] P. Erdős and A. Rényi. On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, 38:343–347, 1961.

[EW77] P. Erdős and Robin J. Wilson. On the chromatic index of almost all graphs. *J. Combinatorial Theory Ser. B*, 23(2–3):255–257, 1977.

[GK73] R. L. Graham and D. J. Kleitman. Increasing paths in edge ordered graphs. *Period. Math. Hungar.*, 3:141–148, 1973. Collection of articles dedicated to the memory of Alfréd Rényi, II.

[GLS99] Anant P. Godbole, Ben Lamorte, and Jessica K. Sklar. If rooks could kill: vertex degrees in random bipartite graphs. In *Combinatorics, graph theory, and algorithms, Vol. I, II (Kalamazoo, MI, 1996)*, pages 445–450. New Issues Press, Kalamazoo, MI, 1999.

[GM09] Catherine Greenhill and Brendan D. McKay. Random dense bipartite graphs and directed graphs with specified degrees. *Random Structures Algorithms*, 35(2):222–249, 2009.

[Hak62] S. L. Hakimi. On realizability of a set of integers as degrees of the vertices of a linear graph. I. *J. Soc. Indust. Appl. Math.*, 10:496–506, 1962.

[Hav55] V. Havel. A remark on the existence of finite graphs [czech]. *Časopis Pest. Mat.*, 80:477–480, 1955.

[HHV09] Ararat Harutyunyan, Paul Horn, and Jacques Verstraete. Independent dominationg sets in graphs of girth five. Preprint, March 2009.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[Jan94] Svante Janson. Orthogonal decompositions and functional limit theorems for random graph statistics. *Mem. Amer. Math. Soc.*, 111(534):vi+78, 1994.

[JŁR00] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[Łuc91] Tomasz Łuczak. A note on the sharp concentration of the chromatic number of random graphs. *Combinatorica*, 11(3):295–297, 1991.

[McD89] Colin McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.

[McD98] Colin McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms Combin.*, pages 195–248. Springer, Berlin, 1998.

[MW90] Brendan D. McKay and Nicholas C. Wormald. Asymptotic enumeration by degree sequence of graphs of high degree. *European J. Combin.*, 11(6):565–580, 1990.

[MW97] Brendan D. McKay and Nicholas C. Wormald. The degree sequence of a random graph. I. The models. *Random Structures Algorithms*, 11(2):97–117, 1997.

[Odl95] A. M. Odlyzko. Asymptotic enumeration methods. In *Handbook of combinatorics, Vol. 1, 2*, pages 1063–1229. Elsevier, Amsterdam, 1995.

[Pal84] Z. Palka. On the degrees of vertices in a bichromatic random graph. *Period. Math. Hungar.*, 15(2):121–126, 1984.

[Pal87] Zbigniew Palka. Extreme degrees in random graphs. *J. Graph Theory*, 11(2):121–134, 1987.

[Pal88] Zbigniew Palka. Asymptotic properties of random graphs. *Dissertationes Math. (Rozprawy Mat.)*, 275:112, 1988.

[Ruc81] Andrzej Ruciński. The r-connectedness of k-partite random graph. *Bull. Acad. Polon. Sci. Sér. Sci. Math.*, 29(7-8):321–330, 1981.

[Rys63] Herbert John Ryser. *Combinatorial mathematics*. The Carus Mathematical Monographs, No. 14. Published by The Mathematical Association of America, 1963.

[SS87] E. Shamir and J. Spencer. Sharp concentration of the chromatic number on random graphs $G_{n,p}$. *Combinatorica*, 7(1):121–129, 1987.

[Vu02] V. H. Vu. Concentration of non-Lipschitz functions and applications. *Random Structures Algorithms*, 20(3):262–316, 2002. Probabilistic methods in combinatorial optimization.

[Wil91] David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.

[Wil06] Herbert S. Wilf. *generatingfunctionology*. A K Peters Ltd., Wellesley, MA, third edition, 2006.

[ZS02] Arif Zaman and Daniel Simberloff. Random binary matrices in biogeographical ecology—instituting a good neighbor policy. *Environ. Ecol. Stat.*, 9(4):405–421, 2002.