An Introduction to the Theory of Random Graphs

Pavan Sangha

July 3, 2014

Contents

Chapter 1

Abstract

This thesis provides an introduction to the fundamentals of random graph theory. The study starts introduces the two fundamental building blocks of random graph theory, namely discrete probability and graph theory. The study starts by introducing relevant concepts probability commonly used in random graph theory- these include concentration inequalities such as Chebyshev's inequality and Chernoff's inequality. Moreover we proceed by introducing central concepts in graph theory, which will underpin the later discussion. In particular we provide results such as Mycielski's construction of a family of triangle-free graphs with high chromatic number and results in Ramsey theory. Next we introduce the concept of a random graph and present two of the most famous proofs in graph theory using the theory random graphs. These include the proof of the fact that there are graphs with arbitrarily high girth and chromatic number, and a bound on the Ramsey number $R(k, k)$. Finally we conclude by introducing the notion of a threshold function for a monotone graph property and we present proofs for the threhold functions of certain properties.

Chapter 2

Preface

Random graph theory has become a major field of study in discrete mathematics and theoretical computer science. Work on the theory of random graphs is extensive but in addition it is used in many applied areas of research, especially in the field of complex networks. A complex network is a graph which often exhibits non-trivial structural features. Examples include biological, airline route, internet and social networks. Take for instance a social network, we can consider the people in the network to be vertices and consider friendship between two people to be represented by an edge. Our social network is likely to be far too large to examine so we resort to looking at approximate models. For example what do we expect the average number of friends of each individual in the network to be? Or how big do we expect the largest group of people all of who know each other to be? Random graphs provide a starting point for many cases where modeling a complex network is required. A detailed treatment of complex networks can be found in [16] and [17]. In addition to providing a framework for modeling complex networks, random graph theory provides us with answers to problems in graph theory. In graph theory we are often interested in the properties of a "typical" graph. For example what is the length of the shortest cycle in a "typical graph"? Or the size of the largest clique in a "typical graph"? A very good way to obtain a measure for such questions is to use random graphs.

Random graph theory is an area of combinatorics which combines both graph theory and probability theory. In the late 1940's, the Hungarian mathematician Paul Erdős realized that probabilistic tools were useful in tackling extremal problems in graph theory. In 1959 Erdős and Réyni collaborated

on a seminal paper [10] that paved the way for the development of the subject now known as random graph theory. In this paper they introduced the uniform random graph model. In this model a graph with n vertices and m edges is selected at random with each such graph having equal probability of being selected. Erdős and Réyni showed that for several natural monotone increasing graph properties, graphs with a number of edges slightly less than a particular threshold were very unlikely to satisfy the property whereas graphs with slightly more edges than the threshold were very likely to satisfy the property. Around the same time the mathematician Edgar Gilbert proposed the now more commonly used binomial random graph model [12]. In the binomial model each potential edge of a graph with n vertices is either selected or not selected with a fixed probability independently of the other potential edges. We will examine both models in this thesis and compare them.

We start this thesis by introducing the fundamentals of probability theory in the third chapter. We first introduce the concept of a probability space and then introduce the concept of a random variable. We end this chapter by studying expectation and variance of a random variable. These two concepts are central to our study of random graphs.

In the fourth chapter we define the binomial distribution and prove important results required for our study of the binomial random graph model. We then give an overview of other important discrete probability distributions and calculate the expectation and variance for each of them. Next we introduce Markov's inequality and Chebyshev's inequality. Both of these inequalities will be used heavily in chapter 6 when studying threshold functions for monotone graph properties. The study of moment generating functions will lead us to prove the substantially better Chernoff bounds.

We look at the fundamentals of graph Theory in chapter five. Basic terminology will be introduced first and then we will focus on a variety of fundamental concepts such as of Paths and Cycles, Connectivity, Trees and Forests and Bipartite graphs. We end this chapter by introducing the chromatic number of a graph and Ramsey theory; topics for which results using random graphs will be provided in chapter 5.

In the sixth chapter we introduce the uniform random graph model and

prove some it's properties. We then compare these with analogous properties obtained from the binomial random graph model. We will also introduce topics such as staged exposure, which allows us to view a binomial random graph as the union of two other binomial random graphs. In addition using random graphs we will show the existence of graphs with arbitrarily large girth and chromatic number. Finally we will show that the uniform random graph model and the binomial random graph model are asymptotically equivalent when the expected number of edges of both models is the same.

We will conclude our thesis with the study of threshold functions for monotone graph properties in chapter seven. The chapter will start with the introduction of threshold functions for particular types of natural classes of graphs, such as trees and small cliques. We will then move on to determine the threshold function for any balanced graph. Finally we will introduce the notion of Sharp threshold functions and we will prove that the property of connectivity has a sharp threshold function.

Chapter 3

Introduction to Probability Theory

3.1 Probability Space

We start by defining the concept of a probability space.

Definition 3.1. A probability space is a triple $(\Omega, \mathcal{F}, P)$ where the following hold.

- Ω is a sample space which is the set of all atom events.
- $\mathcal{F} \subseteq \mathcal{P}(\Omega)$-is a σ-algebra and is called the event space. Furthermore elements of $\mathcal{F}$ are known as events.
- $P : \mathcal{F} \to [0, 1]$ is a probability function which satisfies:
  1. $P(\emptyset) = 0$ and $P(\Omega) = 1$.
  2. $0 \leq P(A) \leq 1$ for all events $A \in \mathcal{F}$.
  3. Let A, B be disjoint events, $A, B \in \mathcal{F}$. Then $P(A \cup B) = P(A) + P(B)$.

For the purposes of this thesis, we will only concern ourselves with Discrete Probability Theory. It therefore suffices to assume that Ω is countable. Moreover we will always assume that $\mathcal{F} = \mathcal{P}(\Omega)$.

Let's start with a simple example of of a probability space.

Example 3.2. Consider tossing a fair coin twice. Our probability space $(\Omega, \mathcal{P}(\Omega), P)$ has sample space

$$\Omega = \{HH, HT, TH, TT\},$$

with $P : \mathcal{P}(\Omega) \to [0, 1]$ defined on the atom events as follows

$$P(HH) = P(HT) = P(TH) = P(TT) = 1/4.$$

Now consider the event A that we see an even number of heads. Clearly

$$A = \{HH, TT\} = \{HH\} \cup \{TT\},$$

hence $P(A) = P(\{HH\}) + P(\{TT\}) = 1/2$ using part 3 of Definition, 3.1.

3.1.1 A review of basic probability

Proposition 3.3. *Let $(\Omega, \mathcal{P}(\Omega), P)$ be a probability space and let $A, B \in \mathcal{P}(\Omega)$ be two events. Then*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Proof. The even A can be expressed as follows

$$(A \setminus B) \cup (A \cap B) = A$$

and furthermore the two sets $(A \setminus B)$ and $(A \cap B)$ are disjoint. Using part 3 of Definition 3.1 it follows that

$$P(A) = P(A \setminus B) + P(A \cap B) \tag{3.1}$$

and similarly

$$P(B) = P(B \setminus A) + P(A \cap B). \tag{3.2}$$

We can represent $A \cup B$ as

$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$$

and as these three sets are disjoint. Using part 3 of Definition 3.1 we obtain the following

$$P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B). \tag{3.3}$$

Finally adding equations (2.1) and (2.2) we obtain

$$P(A) + P(B) = P(A \setminus B) + P(B \setminus A) + 2P(A \cap B)$$

and it follows that

$$\begin{aligned} P(A) + P(B) - P(A \cap B) &= P(A \setminus B) + P(B \setminus A) + P(A \cap B) \\ &= P(A \cup B) \end{aligned}$$

with the last equality following from equation (2.3). □

We now introduce the concept of conditional probability which is essentially the probability of events occurring given that other events have occurred.

Definition 3.4. Let $(\Omega, \mathcal{P}(\Omega), P)$ be a probability space and let $A, B \in \mathcal{P}(\Omega)$ be two events. Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)} ,$$

where $P(A|B)$ is the probability that event A occurs given that event B has occurred.

Example 3.5. Consider the same sample space given in Example 3.2. Let A be the event that the first coin is a head and B be the even that both coins give a head. Then

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}.$$

On the other hand,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{1}{4}} = 1.$$

Definition 3.6. Let $(\Omega, \mathcal{P}(\Omega), P)$ be a probability space and let $A, B \in \mathcal{P}(\Omega)$ be two events. We say that A and B are independent if

$$P(A \cap B) = P(A)P(B).$$

Proposition 3.7. *If A and B are independent then $P(A|B) = P(A)$.*

Proof. it follows that

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(A)P(B)}{P(B)} \\ &= P(A)\ , \end{aligned}$$

with the second equality following from Definition 3.6. □

The notion of conditional probability leads us nicely to our next proposition concerning the law of total probability.

Proposition 3.8. *Let $(\Omega, \mathcal{P}(\Omega), P)$ be a probability space, $\mathcal{I} \subseteq \mathbb{N}$ and let $\{\mathcal{B}_i : i \in \mathcal{I}\}$ be a partition of the sample space Ω, then for every event $A \in \mathcal{P}(\Omega)$:*

$$P(A) = \sum_{i \in \mathcal{I}} P(A|\mathcal{B}_i)P(\mathcal{B}_i).$$

Proof. We can represent A by the following union

$$A = \bigcup_{i \in \mathcal{I}} (A \cap \mathcal{B}_i)$$

as

$$\bigcup_{i \in \mathcal{I}} \mathcal{B}_i = \Omega.$$

Furthermore

$$(A \cap B_i) \cap (A \cap B_j) = \emptyset \quad \text{for all} \quad i \neq j$$

as

$$\mathcal{B}_i \cap \mathcal{B}_j = \emptyset.$$

Hence

$$\begin{aligned} P(A) &= P(\bigcup_{i\in\mathcal{I}}(A \cap \mathcal{B}_i)) \\ &= \sum_{i\in\mathcal{I}} P(A \cap \mathcal{B}_i) \\ &= \sum_{i\in\mathcal{I}} P(A|\mathcal{B}_i)P(\mathcal{B}_i) \end{aligned}$$

with the second equality following from part 3 of Definition 3.1 and the final equality following from Definition 3.4. □

3.1.2 Random Variables

Dealing with sample spaces can become increasingly difficult as the sample space can get extremely large. For example consider tossing a coin 100 times. This sample space denoted Ω contains 2^{100} elements which can be considered a relatively large number. In fact each time we toss a coin once more our sample space doubles in size. We may just be interested in the number of heads which appear in our 100 tosses. It's possible to define a mapping from $X : \Omega \to [100]$ where Ω is the sample space described above. Given $\omega \in \Omega$, $X(\omega)$ is equal to the number of heads which appear in ω. For each $x \in [100]$ let $f(x)$ be the number of ω in Ω with x heads (i.e the number of $\omega \in \Omega$ satisfying $X(\omega) = x$). We could then think of associating a probability with each element in [100], as follows $P(X = x) = \frac{f(x)}{|\Omega|} = \frac{f(x)}{2^{100}}$.

Definition 3.9. Given a probability space $(\Omega, \mathcal{P}(\Omega), P)$, a discrete random variable X is a mapping $X : \Omega \to \Omega_X$ where Ω_X is a finite or countable set (in this thesis $\Omega_X \subseteq \mathbb{Z}$) .

Definition 3.10. The probability mass function (pmf) is a function satisfying $P_X : \Omega_X \to [0, 1]$ (which we will later denote as P for simplicity) and is defined as follows

$$P_X(k) = P_X(X = k) = P(\{\omega \in \Omega : X(\omega) = k\}).$$

It is possible to extend the probability mass function, defined on a random variable to the definition of the probability function defined in Definition 3.1. In order to do this we first introduce some simple ideas.

Proposition 3.11. *Each probability mass function satisfies the following two conditions:*

1. $P_X(k) \geq 0$ *for all* $k \in \Omega_X$.
2. $\sum_{k \in \Omega_X} P_X(k) = 1$.

Proof. Condition 1 follows from Definitions 3.1 and 3.10 since :

$$P_X(k) = P_X(X = k) = P(\{\omega \in \Omega : X(\omega) = k\}) \geq 0.$$

For condition 2, let us start by defining for each $k \in \Omega_X$, the set

$$\Omega_k = \{\omega \in \Omega : X(\omega) = k\}.$$

Each $\omega \in \Omega$ belongs to exactly one such Ω_k and so

$$\Omega = \bigcup_{k \in \Omega_X} \Omega_k$$

and

$$\Omega_k \cap \Omega_l = \emptyset \text{ for any } k \neq l \in \Omega_X$$

Hence,

$$\begin{aligned} P(\Omega) &= P(\cup_{k \in \Omega_X} \Omega_k) \\ &= \sum_{k \in \Omega_X} P(\Omega_k) \\ &= \sum_{k \in \Omega_X} P(\{\omega \in \Omega : X(\omega) = k\}) \\ &= \sum_{k \in \Omega_X} P_X(X = k) \\ &= \sum_{k \in \Omega_X} P_X(k). \end{aligned}$$

with the second equality following from part 3 of Definition 3.1 and the fourth equality following from Definition 3.10. Finally using part 1 of Definition 3.1 it follows that $\sum_{k \in \Omega_X} P_X(k) = P(\Omega) = 1$. □

Theorem 3.12. *Let Ω be a countable set and $p : \Omega \to [0,1]$ with $\sum_{\omega\in\Omega} p(\omega) = 1$. Then if $P : \Omega \to [0,1]$ and*

$$P(\emptyset) = 0, \text{ and } P(A) = \sum_{\omega\in A} p(\omega) \quad \text{for all,} \quad A \in \mathcal{P}(\Omega) ,$$

we claim $(\Omega, \mathcal{P}(\Omega), P)$ is a probability space.

Proof. It suffices to check that our probability function P is consistent with Definition 3.1.

1. $P(\emptyset) = 0$ and $P(\Omega) = \sum_{\omega\in\Omega} p(\omega) = 1$ both hold by definition and our assumption on P and so the first condition holds.
2. The second condition holds because for all $A \in \mathcal{P}(\Omega)$, it follows that

$$\begin{aligned} 0 \le P(A) &= \sum_{\omega\in A} p(\omega) \\ &\le \sum_{\omega\in\Omega} p(\omega) + \sum_{\omega\in\Omega\setminus A} p(\omega) \\ &= \sum_{\omega\in\Omega} p(\omega) \\ &= 1. \end{aligned}$$

3. Let $A, B \in P(\Omega)$ be disjoint events. Then

$$\begin{aligned} P(A \cup B) &= \sum_{\omega\in A\cup B} p(\omega) \\ &= \sum_{\omega\in A} p(\omega) + \sum_{\omega'\in B} p(\omega') \\ &= P(A) + P(B). \end{aligned}$$

Since all three conditions are satisfied it follows that P is a probability function. □

Corollary 3.13. *Given a random variable X and pmf P_X we can define a probability space $(\Omega_X, \mathcal{P}(\Omega_X), P)$.*

Proof. We define

$$P(\emptyset) = 0 \text{ and } P(A) = \sum_{k \in A} P_X(k) \quad \text{for all } A \in \mathcal{P}(X).$$

Then $(\Omega_X, \mathcal{P}(\Omega_X), P)$ is a probability space using Theorem 3.12 and Proposition 3.11. □

Example 3.14. Consider tossing a coin 3 times. Our sample space $\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$. Now consider the set $\Omega_X = \{0, 1, 2, 3\}$ and the random variable $X : \Omega \to \Omega_X$ with X mapping an element of Ω to the number of heads in it. Hence

$$\begin{aligned} X(HHH) &= 3 \\ X(THH) &= X(HTH) = X(HHT) = 2 \\ X(HTT) &= X(THT) = X(TTH) = 1 \\ X(TTT) &= 0. \end{aligned}$$

It follows that,

$$\begin{aligned} P(X = 0) = P_X(X = 0) &= P(\{TTT\}) = \frac{1}{8}, \\ P(X = 1) = P_X(X = 1) &= P(\{HTT, THT, TTH\}) = \frac{3}{8}, \\ P(X = 2) = P_X(X = 2) &= P(\{HHT, HTH, THH\}) = \frac{3}{8}, \\ P(X = 3) = P_X(X = 3) &= P(\{HHH\}) = 1/8. \end{aligned}$$

3.1.3 Jointly distributed random variables

Given two random variables X and Y it is often useful to combine the two and form a joint distribution (X, Y).

Definition 3.15. Given two random variables X and Y we define a new random variable $f(X, Y) : \Omega_X \times \Omega_Y \to \mathbb{Z}$ (often in this thesis $f(X, Y) = X+Y$ or $f(X, Y) = XY$) The probability mass function is defined on $\Omega_X \times \Omega_Y$ and defined by

$$P_{XY} = P(X = x, Y = y), \text{ with } \sum_{(x,y) \in \Omega_X \times \Omega_Y} P(X = x, Y = y) = 1$$

The following are obtain the individual probability mass functions for X, and Y as

$$P_X(X = i) = \sum_{y \in \Omega_Y} P_{XY}(X = i, Y = y)$$

and

$$P_Y(Y = j) = \sum_{x \in \Omega_X} P_{XY}(X = x, Y = j).$$

Theorem 3.16. *Given a random variable with probability mass function P_{XY} where X and Y are random variables. We can define a probability function as follows,*

$$P : \mathcal{P}(\Omega_X \times \Omega_Y) \to [0, 1].$$

Proof. Similar to Theorem 3.12 we define $P(\emptyset) = 0$ and

$$P(A) = \sum_{(\omega_1, \omega_2) \in A} P_{XY}(X = \omega_1, Y = \omega_2) \text{ for all } A \in \mathcal{P}(\Omega_X \times \Omega_Y).$$

Similar to the proof of Theorem 3.12 we can easily verify that this definition is consistent with Definition 3.1. □

Definition 3.17. Given a random variables X and Y we say that X and Y are independent if for all $(x, y) \in \Omega_X \times \Omega_Y$

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y).$$

Example 3.18. Consider a pot with 3 balls numbered $1, 2, 3$, all with equal probability of being selected. Suppose we select two balls without replacement. Let X be the random variable denoting the number of the first ball and Y the random variable denoting the number on the second ball. There are 6 possibilities for the numbers on the first two balls all of which equally likely. Hence

$$P(X = 1, Y = 2) = \frac{1}{6} \neq P(X = 1)P(Y = 2) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9} .$$

Thus X and Y are not independent.

3.1.4 Expectation

When dealing with random variables it is useful to know what kind of values the random variable is likely to take. It is therefore sensible when dealing with probability to formalize the concept of expectation.

Definition 3.19. Given a probability space $(\Omega, \mathcal{P}(\Omega), P)$ and a random variable $X : \Omega \to \Omega_X$ such that $\sum_{k\in\Omega_x} |k|P(X = k) < \infty$ the expectation of X denoted $\mathbb{E}(X)$ is defined as

$$\mathbb{E}(X) = \sum_{k\in\Omega_x} kP(X = k).$$

.

The value $\mathbb{E}(X)$ takes need not be an element of Ω_X itself, while this may seem undesirable the expectation of a random variable provides us with lots of useful information.

Example 3.20. Consider rolling a fair dice. Let the random variable X be the random variable which denote the value on the dice. Hence for any $k \in \Omega_X = \{1, 2, 3, 4, 5, 6\}$ we find $P(X = k) = \frac{1}{6}$. Then

$$\mathbb{E}(X) = \frac{\sum_{i=1}^{6} i}{6} = 3.5$$

Theorem 3.21. *Let X, Y be discrete random variables and let*

$$f(X, Y) = aX + bY.$$

Then it follows that

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

Theorem 3.21 is known as the linearity of expectation and will be used throughout this thesis. One of the reasons the concept is important is because it holds when the random variable X_i are dependent as well as independent.

Proof.

$$\begin{aligned}
\mathbb{E}(aX+bY) &= \sum_{(x,y)\in\Omega_X\times\Omega_Y} (ax+by)P(X=x,Y=y) \\
&= \sum_{(x,y)\in\Omega_X\times\Omega_Y} axP(X=x,Y=y) \\
&+ \sum_{(x,y)\in\Omega_X\times\Omega_Y} byP(X=x,Y=y) \\
&= \sum_{x\in\Omega_X} ax \sum_{y\in\Omega_Y} P(X=x,Y=y) \\
&+ \sum_{y\in\Omega_Y} by \sum_{x\in\Omega_X} P(X=x,Y=y) \\
&= a\sum_{x\in\Omega_X} x \sum_{y\in\Omega_Y} P(X=x,Y=y) \\
&+ b\sum_{y\in\Omega_Y} y \sum_{x\in\Omega_X} P(X=x,Y=y) \\
&= a\sum_{x\in\Omega_X} xP(X=x) \\
&+ b\sum_{y\in\Omega_Y} yP(Y=y) \\
&= a\mathbb{E}(X)+b\mathbb{E}(Y).
\end{aligned}$$

The fifth inequality follows from Definition 3.15. □

Corollary 3.22. *Let $X_1, X_2, \cdots, X_n$ be Discrete random variables and $X = \sum_{i=1}^n a_i X_i$ then*

$$\mathbb{E}(X) = \sum_{i=1}^n a_i \mathbb{E}(X_i).$$

The proof follows easily by induction on n.

Theorem 3.23. *Let X and Y be independent random variables. Then*

$$\mathbb{E}(XY) = \mathbb{E}(X)\cdot\mathbb{E}(Y).$$

Proof.

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_{(x,y)\in\Omega_X\times\Omega_Y} xyP(X=x,Y=y) \\
&= \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} xyP(X=x,Y=y) \\
&= \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} xyP(X=x)P(Y=y) \\
&= \sum_{x\in\Omega_X} xP(X=x)\sum_{y\in\Omega_Y} yP(Y=y) \\
&= \sum_{x\in\Omega_X} xP(X=x)\mathbb{E}(Y) \\
&= \mathbb{E}(Y)\sum_{x\in\Omega_X} xP(X=x) \\
&= \mathbb{E}(Y)\mathbb{E}(X)\,.
\end{aligned}$$

The third equality follows from Definition 3.17 and the definition of independence. □

Example 3.24. We consider a classic example used in theoretical computer science, which shows the usefulness of the linearity of expectation. Consider m balls labeled $1,\ldots,m$ and n bins labeled $1,\ldots,n$. Each ball is then placed into a bin independently and uniformly at random.

Let X_j be the number of balls in bin j once all balls have been placed in a bin. Then it follows that for $j = 1,\ldots,n$

$$\mathbb{E}(X_j) = \frac{m}{n}.$$

Proof. For every $1 \le i \le m$ and $1 \le j \le n$ define an indicator random variable as follows $X_{ij} = 1$ if ball i is placed into bin j and $X_{ij} = 0$ otherwise. Hence

$$X_j = \sum_{i=1}^{m} X_{ij}.$$

Since ball i chooses bin j uniformly at random it follows that

$$P(X_{ij} = 1) = \frac{1}{n}$$

and so

$$\mathbb{E}(X_{ij}) = 1 \cdot P(X_{ij} = 1) + 0 \cdot P(X_{ij} = 0) = \frac{1}{n}.$$

Finally

$$\mathbb{E}(X_j) = \mathbb{E}\left(\sum_{i=1}^{m} X_{ij}\right) = \sum_{i=1}^{m} \mathbb{E}(X_{ij}) = m \cdot \frac{1}{n}.$$

□

3.1.5 Variance

Consider the random variables X_1, X_2 and let the random variable $X = \frac{X_1+X_2}{2}$. Where X_1 and X_2 are the values obtained by rolling two fair dice. Clearly $\mathbb{E}(X_1) = \mathbb{E}(X_2)$ and so by the linearity of expectation

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}\left(\frac{X_1 + X_2}{2}\right) \\
&= \mathbb{E}\left(\frac{X_1}{2}\right) + \mathbb{E}\left(\frac{X_2}{2}\right) \\
&= \frac{1}{2} \cdot \mathbb{E}(X_1) + \frac{1}{2} \cdot \mathbb{E}(X_2) \\
&= \mathbb{E}(X_1).
\end{aligned}$$

Both random variables X_1 and X have the same expectation but their probability distributions are different. The probability distribution of X is more 'bell-shaped' while the distribution of X_1 is uniform. We now introduce a quantity known as variance, which measures how far away from the mean we expect the random variable to be.

Definition 3.25. Let X be a random variable, then

$$\text{var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2.$$

Proposition 3.26. $\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

Proof.

$$\begin{aligned}
\text{var}(X) &= \mathbb{E}(X - \mathbb{E}(X))^2 \\
&= \mathbb{E}(X^2 - 2\mathbb{E}(X)X + \mathbb{E}(X)^2) \\
&= \mathbb{E}(X^2) - \mathbb{E}(2\mathbb{E}(X)X) + \mathbb{E}(\mathbb{E}(X)^2) \\
&= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 \\
&= \mathbb{E}(X^2) - \mathbb{E}(X)^2.
\end{aligned}$$

The third and fourth inequalities follow from Theorem 3.21 (linearity of expectation) □

Example 3.27. Let us go back to the scenario introduced at the start of the section. Let X_1 and X_2 be the random variables who's values are obtained by rolling two fair dice and X be the random variable whose value is the average of X_1 and X_2. As mentioned before using the linearity of expectation it follows that X and X_1 have the same expectation but have different probability distributions. We now look at how the variance of the random variables X and X_1 differ.
We start by computing the $\mathbb{E}(X_1^2)$ it follows that,

$$\begin{aligned}\mathbb{E}(X_1^2) &= \sum_{i=1}^{6} i^2 \cdot \frac{1}{6} \\ &= \frac{1}{6}\sum_{i=1}^{6} i^2 \\ &= 15.1667 \text{ (4dp)}.\end{aligned}$$

Using Proposition 3.26,

$$\begin{aligned}\text{var}(X_1) &= \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2 \\ &\approx 15.1667 - 3.5^2 \\ &= 2.1967\ .\end{aligned}$$

Next we compute $\mathbb{E}(X^2)$, it follows that

$$\begin{aligned}\mathbb{E}(X^2) &= 1^2\frac{1}{36} + 1.5^2\frac{2}{36} + 2^2\frac{3}{36} + 2.5^2\frac{4}{36} + 3^2\frac{5}{36} + 3.5^2\frac{6}{26} \\ &+ 4^2\frac{5}{36} + 4.5^2\frac{4}{36} + 5^2\frac{3}{36} + 5.5^2\frac{2}{36} + 6^2\frac{1}{36} \\ &= 13.7083 \text{ 4dp.}\end{aligned}$$

Using Proposition 3.26, we obtain that

$$\begin{aligned}\text{var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ &\approx 13.7083 - 3.5^2 \\ &= 1.4583\ .\end{aligned}$$

We can see that as the variance of X is less than the variance of X_1 we expect X to be closer to it's expected value.

Theorem 3.28. *Let X be a random variable and a and b constants. Then*

$$\text{var}(aX + b) = a^2\text{var}(X).$$

Proof.

$$\begin{aligned}
\text{var}(aX+b) &= \mathbb{E}((aX+b)^2) - \mathbb{E}(aX+b)^2 \\
&= \mathbb{E}(a^2X^2 + 2abX + b^2) - (a\mathbb{E}(X) + b)^2 \\
&= (a^2\mathbb{E}(X^2) + 2ab\mathbb{E}(X) + b^2) - (a^2\mathbb{E}(X)^2 + 2ab\mathbb{E}(X) + b^2) \\
&= a^2\mathbb{E}(X^2) - a^2\mathbb{E}(X)^2 \\
&= a^2(\mathbb{E}(X^2) - \mathbb{E}(X)^2) \\
&= a^2\text{var}(X).
\end{aligned}$$

The first equality follows from Proposition 3.26 and the third and fourth equality's follow from Theorem 3.21 (linearity of expectation). □

Definition 3.29. If X and Y are random variables then the covariance of X and Y is defined as

$$\text{cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))).$$

Theorem 3.30. *Let X and Y be random variables then*

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Proof.

$$\begin{aligned}
\text{cov}(X,Y) &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \\
&= \mathbb{E}(XY - \mathbb{E}(X)Y - \mathbb{E}(Y)X + \mathbb{E}(X)\mathbb{E}(Y)) \\
&= \mathbb{E}(XY) - \mathbb{E}(\mathbb{E}(X)Y) - \mathbb{E}(\mathbb{E}(Y)X) + \mathbb{E}(\mathbb{E}(X)\mathbb{E}(Y)) \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(Y)\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y) \\
&= \mathbb{E}(XY) - 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned}$$

The third and fourth equality's follow from the linearity of expectation. □

Corollary 3.31 (Symmetry)**.** *Let X and Y be random variables. Then*

$$\text{cov}(X, Y) = \text{cov}(Y, X).$$

Proof.

$$\begin{aligned} \text{cov}(X, Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\ &= \mathbb{E}(YX) - \mathbb{E}(Y)\mathbb{E}(X) \\ &= \text{cov}(Y, X). \end{aligned}$$

The first equality follows from Theorem 3.30. □

Corollary 3.32. *Let X and Y be random variables. Then*

$$\text{var}(X) = \text{cov}(X, X).$$

Proof.

$$\begin{aligned} \text{cov}(X, X) &= \mathbb{E}(X \cdot X) - \mathbb{E}(X) \cdot \mathbb{E}(X) \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ &= \text{var}(X) \end{aligned}$$

The first equality follows from Theorem 3.30. □

Corollary 3.33. *Let X and Y be random variables and a be a constant. Then*

$$\text{cov}(aX, Y) = \text{cov}(X, aY) = a\text{cor}(X, Y).$$

Proof.

$$\begin{aligned} \text{cov}(aX, Y) &= \mathbb{E}(aXY) - \mathbb{E}(aX)\mathbb{E}(Y) \\ &= a\mathbb{E}(XY) - a\mathbb{E}(X)\mathbb{E}(Y) \\ &= \mathbb{E}(XaY) - \mathbb{E}(X)\mathbb{E}(aY) \\ &= \text{cov}(X, aY). \end{aligned}$$

Factoring the second equality gives

$$\begin{aligned}\operatorname{cov}(aX, Y) &= a\mathbb{E}(XY) - a\mathbb{E}(X)\mathbb{E}(Y) \\ &= a(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)) \\ &= a \cdot \operatorname{cov}(X, Y).\end{aligned}$$

The first equality following from Theorem 3.30 and the second and third from the linearity of expectation. □

Corollary 3.34 (Bi-linearity: part a)**.** *Let X, Y and Z be random variables then*

$$\operatorname{cov}(X + Z, Y) = \operatorname{cov}(X, Y) + \operatorname{cov}(Z, Y).$$

Proof.

$$\begin{aligned}\operatorname{cov}(X + Z, Y) &= \mathbb{E}((X + Z)Y) - \mathbb{E}(X + Z)\mathbb{E}(Y) \\ &= \mathbb{E}(XY + ZY) - \mathbb{E}(X + Z)\mathbb{E}(Y) \\ &= \mathbb{E}(XY) + \mathbb{E}(ZY) - (\mathbb{E}(X) + \mathbb{E}(Z))\mathbb{E}(Y) \\ &= \mathbb{E}(XY) + \mathbb{E}(ZY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(Z)\mathbb{E}(Y) \\ &= (\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)) + (\mathbb{E}(ZY) - \mathbb{E}(Z)\mathbb{E}(Y)) \\ &= \operatorname{cov}(X, Y) + \operatorname{cov}(Z, Y)\end{aligned}$$

and the result follows. □

Corollary 3.35 (Bi-linearity:part b)**.** *Let X and Y and Z be a random variables then*

$$\operatorname{cov}(X, Y + Z) = \operatorname{cov}(X, Y) + \operatorname{cov}(X, Z).$$

The proof is similar to Bi-linearity:part a.

Corollary 3.36. *Let X and Y be random variables and c and b constants. Then*

$$\operatorname{cov}(X + a, Y + b) = \operatorname{cov}(X, Y)$$

Proof.

$$\begin{aligned}
\operatorname{cov}(X+a, Y+b) &= \mathbb{E}((X+a)(Y+b)) - \mathbb{E}(X+a)\mathbb{E}(Y+b) \\
&= \mathbb{E}(XY + aY + bX + ab) \\
&- (\mathbb{E}(X) + \mathbb{E}(a))(\mathbb{E}(Y) + \mathbb{E}(b)) \\
&= (\mathbb{E}(XY) + \mathbb{E}(aY) + \mathbb{E}(bX) + \mathbb{E}(ab)) \\
&- (\mathbb{E}(X) + a)(\mathbb{E}(Y) + b) \\
&= (\mathbb{E}(XY) + a\mathbb{E}(Y) + b\mathbb{E}(X) + ab) \\
&- (\mathbb{E}(X)\mathbb{E}(Y) + a\mathbb{E}(Y) + b\mathbb{E}(X) + ab) \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\
&= \operatorname{cov}(X, Y).
\end{aligned}$$

The first and final equality's follows from Theorem 3.30 and the second,third and fourth from the linearity of expectation. □

Theorem 3.37. *Let $X = \sum_{i=1}^{n} X_i$ and $Y = \sum_{j=1}^{m} Y_j$ be random variables. Then*

$$\begin{aligned}
\operatorname{cov}(X, Y) &= \operatorname{cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \operatorname{cov}(X_i, Y_j).
\end{aligned}$$

Proof.

$$\begin{aligned}
\operatorname{cov}(X,Y) &= \mathbb{E}\left(\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{j=1}^{m} Y_j\right)\right) - \mathbb{E}\left(\sum_{i=1}^{n} X_i\right)\mathbb{E}\left(\sum_{j=1}^{m} Y_j\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{m} X_i Y_j\right) - \mathbb{E}\left(\sum_{i=1}^{n} X_i\right)\mathbb{E}\left(\sum_{j=1}^{m} Y_j\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbb{E}(X_i Y_j) - \left(\sum_{i=1}^{n} \mathbb{E}(X_i)\right)\left(\sum_{j=1}^{m} \mathbb{E}(Y_j)\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbb{E}(X_i Y_j) - \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbb{E}(X_i)\mathbb{E}(Y_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbb{E}(X_i Y_j) - \mathbb{E}(X_i)\mathbb{E}(Y_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \operatorname{cov}(X_i, Y_j).
\end{aligned}$$

The first and final equalities follow from Theorem 3.30 and the third and fourth equalities from the linearity of expectation. □

Theorem 3.38. *Let $X_1, X_2, \ldots, X_n$ be independent random variables, then*

$$\operatorname{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{var}(X_i)$$

Proof. We may assume $\mathbb{E}(X_i) = 0$ for each i, otherwise we may replace each X_i with $X_i - \mathbb{E}(X_i)$ because it follows that

$$\operatorname{var}(X_1 + X_2 + \ldots + X_n) = \operatorname{var}(X_1 - \mathbb{E}(X_1) + X_2 - \mathbb{E}(X_2) + \ldots + X_n - \mathbb{E}(X_n))$$

using Theorem 3.28 we obtain that

$$\begin{aligned}
\operatorname{var}(X_i) &= \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 \\
&= \mathbb{E}(X_i^2) - 0^2 \\
&= \mathbb{E}(X_i^2).
\end{aligned}$$

Furthermore it follows from the linearity of expectation that

$$\mathbb{E}(X_1 + X_2 + \ldots + X_n) = 0$$

and

$$\begin{aligned}\mathbb{E}(X_i X_j) &= \mathbb{E}(X_i)\mathbb{E}(X_j) \\ &= 0 \cdot 0 \\ &= 0\end{aligned}$$

using the fact that the X_i, X_j are independent.
Finally

$$\begin{aligned}\mathrm{var}(X_1 + X_2 + \ldots + X_n) &= \mathbb{E}((X_1 + X_2 + \ldots + X_n)^2) - \mathbb{E}(X_1 + X_2 + \ldots + X_n)^2 \\ &= \mathbb{E}((X_1 + X_2 + \ldots + X_n)^2) - 0^2 \\ &= \mathbb{E}((X_1 + X_2 + \ldots + X_n)^2) \\ &= \mathbb{E}\left(\sum_{1 \le i,j \le n} X_i X_j\right) \\ &= \sum_{1 \le i,j \le n} \mathbb{E}(X_i X_j) \\ &= \sum_{1 \le i \le n} \mathbb{E}(X_i^2) + \sum_{1 \le i \ne j \le n} \mathbb{E}(X_i X_j) \\ &= \sum_{1 \le i \le n} \mathbb{E}(X_i^2) + \sum_{1 \le i \ne j \le n} 0 \\ &= \sum_{1 \le i \le n} \mathbb{E}(X_i^2) \\ &= \sum_{1 \le i \le n} \mathrm{var}(X_i) \\ &= \sum_{i=1}^{n} \mathrm{var}(X_i)\end{aligned}$$

and the result follows. □

Theorem 3.39. *Let $X_1, X_2, \ldots, X_n$ be independent random variables, with $\mathrm{var}(X_i) = \sigma_i^2$. Then*

$$\mathrm{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{var}(X_i) + 2\sum_{i<j} \mathrm{cov}(X_i, X_j).$$

Proof.

$$\begin{aligned}
\mathrm{var}(X) &= \mathbb{E}\left((X_1 + X_2 + \ldots + X_n)^2\right) - \left(\mathbb{E}(X_1 + X_2 + \ldots + X_n)\right)^2 \\
&= \mathbb{E}\left(\sum_{i=1}^{n} X_i^2 + 2\sum_{i<j} X_i X_j\right) - \left(\mathbb{E}(X_1) + \mathbb{E}(X_2) + \ldots + \mathbb{E}(X_n)\right))^2 \\
&= \mathbb{E}\left(\sum_{i=1}^{n} X_i^2 + 2\sum_{i<j} X_i X_j\right) - \left(\sum_{i=1}^{n} \mathbb{E}(X_i)^2 + 2\sum_{i<j} \mathbb{E}(X_i)\mathbb{E}(X_j)\right) \\
&= \left(\sum_{i=1}^{n} \mathbb{E}(X_i^2)\right) + 2\left(\sum_{i<j} \mathbb{E}(X_i X_j)\right) - \sum_{i=1}^{n} \mathbb{E}(X_i)^2 - 2\left(\sum_{i<j} \mathbb{E}(X_i)\mathbb{E}(X_j)\right) \\
&= \left(\sum_{i=1}^{n} \mathbb{E}(X_i^2) - \sum_{i=1}^{n} \mathbb{E}(X_i)^2\right) + 2\left(\sum_{i<j} \mathbb{E}(X_i X_j) - \mathbb{E}(X_i X_j)\right) \\
&= \left(\sum_{i=1}^{n} \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2\right) + 2\left(\sum_{i<j} \mathbb{E}(X_i X_j) - \mathbb{E}(X_i X_j)\right) \\
&= \sum_{i=1}^{n} \mathrm{var}(X_i) + 2\sum_{i<j} \mathrm{cov}(X_i, X_j).
\end{aligned}$$

the first equality follows from proposition 3.26 and the final equality follows from proposition 3.26 and theorem 3.30. □

Chapter 4

Further Probability Theory

4.1 Discrete probability distributions

We have introduced the notions of a probability space and a Random variable in the previous chapter. In this chapter we look at different types of discrete probability distributions, with the binomial distribution being central to our later work. Given a random variable X a discrete probability distribution assigns a probability to each value that X can take. We start by taking a look at arguably the simplest probability distribution known as a Bernoulli trial.

Definition 4.1. Given a probability space $(\Omega, \mathcal{P}(\Omega), P)$ where $\Omega = \{x, y\}$, let X be a random variable $X : \Omega \to \Omega_X$ where $\Omega_X = \{0, 1\}$. Furthermore let $P(X = 1) = p$ and $P(X = 0) = 1 - p$ where $0 \leq p \leq 1$. We then say that X is a Bernoulli trial.

Example 4.2. A simple example illustrating a Bernoulli trial, would be tossing a biased dice with probability $0 \leq p \leq 1$ of observing a head. If we observe a head our random variable takes the value 1 and if it is a tail it takes the value 0.

Theorem 4.3. *Let Y be a Bernoulli trial. Then $\mathbb{E}(Y) = p$ and $\text{var}(Y) = p(1 - p)$.*

Proof. $\mathbb{E}(Y) = 1 \cdot p + 0 \cdot (1 - p) = p$. Furthermore

$$\mathbb{E}(Y^2) = p \cdot 1^2 + (1 - p) \cdot 0^2 = p.$$

Hence,

$$\begin{aligned} \text{var}(Y) &= \mathbb{E}(Y^2) - \mathbb{E}^2(Y) \\ &= p - p^2 \\ &= p(1-p). \end{aligned}$$

and the result follows. □

4.1.1 Binomial Distribution

The Binomial distribution will play a key role in our later work with Binomial Random graphs. We use a motivating example first to introduce the concept.

Example 4.4. Consider tossing a biased coin 3 times. Consider the random variable X_i which is assigned a 1 if we observe H (a head) and 0 if we observe T (a tail) on toss $1 \leq i \leq 3$. Let the random variable $X = \sum_{i=1}^{3} X_i$ (note X counts the total number of 1's). We can represent the outcome of our tosses by the 3-tuple (X_1, X_2, X_3). Thus obtaining HTT would correspond to the 3-tuple $(1, 0, 0)$. There are 8 possible outcomes for the 3-tuple (X_1, X_2, X_3) corresponding to 4 possible outcomes for X. Let's now count these. The 3-tuple $(0, 0, 0)$ corresponds to $X = 0$, each 3-tuple in the set $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ corresponds to $X = 1$, each 3-tuple in the set $\{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}$ corresponds to $X = 2$ and finally $(1, 1, 1)$ corresponds to $X = 3$. Recall that $P(X_i = 1) = p$, hence it follows that the probability that our outcome is the 3-tuple $(0, 0, 0)$ is $(1-p)^3$. Each 3-tuple in the set $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ has a probability of $p(1-p)^2$ of being our outcome. Similarly each 3-tuple in the set $\{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}$ has a probability of $p^2(1-p)$ of being our outcome and finally the 3-tuple $(1, 1, 1)$ has a probability p^3 of being our outcome. We can conclude that

$$\begin{aligned} P(X = 0) &= (1-p)^3 \\ P(X = 1) &= 3p(1-p)^2 \\ P(X = 2) &= 3p^2(1-p) \\ P(X = 3) &= p^3. \end{aligned}$$

Note that

$$\begin{aligned}\sum_{i=1}^{3} P(X_i = 1) &= (1-p)^3 + 3p(1-p)^2 + 3p^2(1-p) + p^3 \\ &= (p + (1-p))^3 \\ &= 1\end{aligned}$$

and so it follows that P is a probability function.

It is possible for us to obtain a generalized formula for $P(X = i)$ for $1 \leq i \leq 3$. For $X = i$ our 3-tuple (X_1, X_2, X_3) will contain exactly i 1's. Thus there are precisely $\binom{3}{i}$ 3-tuples which correspond $X = i$. Furthermore since each of these 3-tuples has a probability of $p^i(1-p)^{3-i}$ of being an outcome we conclude that $P(X = i) = \binom{n}{i}p^3(1-p)^{3-i}$.

Now suppose that we conduct n independent Bernoulli trials and let the random variable X count the number of times we see a 1. We can extend the idea used in the above example to obtain the following generalized formula for $1 \leq i \leq n$

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

We now introduce the concept of the Binomial Distribution.

Definition 4.5. Conduct n independent Bernoulli trials $X_1, X_2, \ldots, X_n$ with $P(X_i = 1) = p$ for $1 \leq i \leq n$. Let X be a random variable counting the number of times we see a 1, thus $X = \sum_{i=1}^{n} X_i$. We say that X is binomially distributed with parameters n and p commonly expressed as $X \sim \text{Bin}(n, p)$.

As mentioned above it follows that $P(X = i) = \binom{n}{i}p^i(1-p)^{n-i}$ for $1 \leq i \leq n$.

Theorem 4.6. *Let $X \sim \text{Bin}(n, p)$ then $\mathbb{E}(X) = np$.*

Proof. Since $X = \sum_{i=1}^{n} X_i$ it follows that

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) \\ &= \sum_{i=1}^{n} \mathbb{E}(X_i) \\ &= \sum_{i=1}^{n} p \\ &= np. \end{aligned}$$

with the second equality following from the linearity of expectation. □

Theorem 4.7. *Let* $X \sim \text{Bin}(n, p)$ *then* $\text{var}(X) = np(1-p)$.

Proof. Since $X = \sum_{i=1}^{n} X_i$ it follows that

$$\begin{aligned} \text{var}(X) &= \text{var}\left(\sum_{i=1}^{n} X_i\right) \\ &= \sum_{i=1}^{n} \text{var}(X_i) \\ &= \sum_{i=1}^{n} p(1-p) \\ &= np(1-p) \end{aligned}$$

with the second inequality following form Theorem 3.38 □

4.1.2 Poisson Distribution

The Poisson distribution is a distribution which counts the number of events X over a period of time, given the average or expected value λ.

Definition 4.8. Let X be a discrete random variable with $\Omega_X = \mathbb{N} \cup \{0\}$. When the probability function $P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$ we say that X is a Poisson random variable with parameter λ.

Theorem 4.9. *If X is a Poisson random variable with parameter λ, then $\mathbb{E}(X) = \lambda$.*

Proof. From Definition 4.8 it follows that

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k\geq 0} k \cdot \frac{1}{k!}\lambda^k e^{-\lambda} \\
&= \sum_{k\geq 1} k \cdot \frac{1}{k!}\lambda^k e^{-\lambda} \\
&= \lambda e^{-\lambda} \sum_{k\geq 1} \frac{1}{(k-1)!}\lambda^{k-1} \\
&= \lambda e^{-\lambda} \sum_{j\geq 0} \frac{\lambda^j}{j!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda.
\end{aligned}$$

The fourth equality follows by setting $j = k - 1$ and the fifth equality follows form the Taylor series expansion of e^{λ}. □

Theorem 4.10. *Let X be a Poisson random variable with parameter λ, then $\mathrm{var}(X) = \lambda$.*

Proof. It follows from Definition 3.26 that $\mathrm{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$. Thus

we start by calculating $\mathbb{E}(X^2)$,

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k\geq 0} k^2 \cdot \frac{1}{k!}\lambda^k e^{-\lambda} \\
&= \sum_{k\geq 1} k^2 \cdot \frac{1}{k!}\lambda^k e^{-\lambda} \\
&= \sum_{k\geq 1} k \cdot \frac{1}{(k-1)!}\lambda^k e^{-\lambda} \\
&= \lambda e^{-\lambda}\left(\sum_{k\geq 1}(k-1)\cdot\frac{1}{(k-1)!}\lambda^{k-1} + \sum_{k\geq 1}\frac{1}{(k-1)!}\lambda^{k-1}\right) \\
&= \lambda e^{-\lambda}\left(\lambda\sum_{k\geq 2}\frac{1}{(k-2)!}\lambda^{k-2} + \sum_{k\geq 1}\frac{1}{(k-1)!}\lambda^{k-1}\right) \\
&= \lambda e^{-\lambda}\left(\lambda\sum_{i\geq 0}\frac{1}{i!}\lambda^{i} + \sum_{j\geq 0}\frac{1}{j!}\lambda^{j}\right) \\
&= \lambda e^{-\lambda}(\lambda e^{\lambda} + e^{\lambda}) \\
&= \lambda(\lambda+1) \\
&= \lambda^2 + \lambda.
\end{aligned}$$

the sixth equality follows by setting $i = k-2$ and $j = k-1$.
Hence we conclude that

$$\begin{aligned}
\text{var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda.
\end{aligned}$$

and the result follows. □

Example 4.11. Suppose we know that on average there is 1 weed per 1m^2 patch of grass in a large park. What is the probability that in a particular 1m^2 patch of grass we see at least 2 weeds?

Let X be the random variable representing the number of weeds in the patch of grass. Then X is a Poisson random variable with parameter $\lambda = 1$. Hence

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - \frac{1^0}{0!}e^{-1} - \frac{1^1}{1!}e^{-1} \approx 0.2642.$$

The Poisson distribution with $\lambda = np$ often provides a very good approximation to the binomial distribution when n is large and p is small.

Theorem 4.12. *Let λ be a constant, let n be a sufficiently large integer and let $p = \frac{\lambda}{n}$. Then for every fixed k*

$$P(X = k) \approx \frac{\lambda^k}{k!}e^{-\lambda}.$$

Proof. Since $\lambda = np$ constant and $X \sim \text{Bin}(n, p)$ it follows that as $n \to \infty$ we have

$$\begin{aligned} P(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} \cdot \frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &\approx 1 \cdot 1 \cdots 1 \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} \cdot 1 \\ &= \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

and the result follows. □

We have used the following two results without proof to obtain the approximation from the third equality:

1. $\lim_{n\to\infty}(1 - \frac{\lambda}{n})^n = e^{-\lambda}$.
2. $\lim_{n\to\infty}(1 - \frac{\lambda}{n})^{-k} = 1$.

4.1.3 Hypergeometric Distribution

We introduce the concept of the Hypergeometric Distribution by starting with a simple example.

Example 4.13. Suppose we have a bag with 5 labeled white balls and 5 labeled black balls. Further more suppose we consider drawing a white ball as a success and drawing a black ball as a failure. If 3 balls are drawn from the bag, let the random variable X count the number of successes. We wish to find a formula for $P(X = i)$ for $i = 0, \ldots, 3$. There are $\binom{10}{3}$ possible choices for our 3 balls. There are $\binom{5}{i}$ ways to obtain i successes and $\binom{5}{3-i}$ to obtain the remaining failures. Hence

$$P(X = i) = \frac{\binom{5}{i} \cdot \binom{5}{3-i}}{\binom{10}{3}}.$$

We now extend the ideas used in the above example to introduce the Hypergeometric Distribution. We first introduce the parameters of our Hypergeometric model.

1. M is the population size
2. K is the total number of possible successes
3. n is the number of samples
4. k is the number of successes in the n samples
5. $\max(0, n - (M - K)) \leq k \leq \min(K, n)$

Definition 4.14. A random variable X follows the Hypergeometric distribution if its probability function is given by

$$P(X = k) = \frac{\binom{K}{k} \cdot \binom{M-K}{n-k}}{\binom{M}{n}}.$$

Theorem 4.15. *Let X be a Hypergeometric random variable, then*

$$\mathbb{E}(X) = \frac{nK}{M}.$$

Proof. It follows from Definition 4.14 that,

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x=0}^{n} \frac{x\binom{K}{x} \cdot \binom{M-K}{n-x}}{\binom{M}{n}} \\
&= \sum_{x=1}^{n} \frac{x\binom{K}{x} \cdot \binom{M-K}{n-x}}{\binom{M}{n}} \\
&= \sum_{x=1}^{n} \frac{x \cdot \frac{K}{x}\binom{K-1}{x-1} \cdot \binom{(M-1)-(K-1)}{(n-1)-(x-1)}}{\frac{M}{n}\binom{M-1}{n-1}} \\
&= \frac{nK}{M} \sum_{x=1}^{n} \frac{\binom{K-1}{x-1} \cdot \binom{(M-1)-(K-1)}{(n-1)-(x-1)}}{\binom{M-1}{n-1}} \\
&= \frac{nK}{M} \sum_{l=0}^{n-1} \frac{\binom{K-1}{l} \cdot \binom{(M-1)-(K-1)}{(n-1)-l}}{\binom{M-1}{n-1}} \\
&= \frac{nK}{M}.
\end{aligned}$$

The 5th equality follows from using the substitution $l = x - 1$ and the last equality follows because the sum of the terms in the fifth equality equal 1. □

Theorem 4.16. *Let X be a Hypergeometric random variable, then*

$$\text{var}(X) = n\frac{K}{M}\left(1 - \frac{K}{M}\right)\frac{M-n}{M-1}.$$

Proof. Recall from Definition 3.25 that $\text{var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2$. Hence it

follows that

$$\begin{aligned}
\text{var}(X) &= \sum_{x=0}^{n} \left(x - \frac{nK}{M}\right)^2 \frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} \\
&= \sum_{x=0}^{n} \frac{x^2\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} - \frac{2nK}{M}\sum_{x=0}^{n} \frac{x\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} + \frac{n^2K^2}{M^2}\sum_{x=0}^{n} \frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} \\
&= \sum_{x=0}^{n} \frac{x^2\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} - \frac{2nK}{M}\mathbb{E}(X) + \frac{n^2K^2}{M^2} \cdot 1 \\
&= \sum_{x=0}^{n} \frac{x^2\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} - \frac{n^2K^2}{M^2}.
\end{aligned}$$

The third equality follows from the fact that $\frac{n^2K^2}{M^2}\sum_{x=0}^{n} \frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} = 1$, and the fourth equality follows from Theorem 4.15.

Next we calculate $\sum_{x=0}^{n} \frac{x^2\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}}$. It follows that

$$\begin{aligned}
\sum_{x=0}^{n} \frac{x^2\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} &= \sum_{x=1}^{n} \frac{x^2\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} \\
&= \sum_{x=1}^{n} \frac{x^2 \cdot \frac{K}{x}\binom{K-1}{x-1}\binom{M-K}{n-x}}{\frac{M}{n}\binom{M-1}{n-1}} \\
&= \frac{nK}{M}\sum_{x=1}^{n} \frac{x\binom{K-1}{x-1}\binom{M-K}{n-x}}{\binom{M-1}{n-1}} \\
&= \frac{nK}{M}\sum_{x=1}^{n} \frac{(x-1)\binom{K-1}{x-1}\binom{M-K}{n-x}}{\binom{M-1}{n-1}} + \frac{nK}{M}\sum_{x=1}^{n} \frac{\binom{K-1}{x-1}\binom{M-K}{n-x}}{\binom{M-1}{n-1}} \\
&= \frac{nK}{M}\sum_{l=0}^{n-1} \frac{l\binom{K-1}{l}\binom{M-K}{(n-1)-l}}{\binom{M-1}{n-1}} + \frac{nK}{M}\sum_{l=0}^{n-1} \frac{\binom{K-1}{l}\binom{M-K}{(n-1)-l}}{\binom{M-1}{n-1}} \\
&= \frac{nK}{M} \cdot \frac{(n-1)(K-1)}{M-1} + \frac{nK}{M} \cdot 1.
\end{aligned}$$

The fourth equality follows from setting $l = x - 1$. The fifth equality follows

from Theorem 4.15 and the fact that $\sum_{l=0}^{n-1} \frac{\binom{K-1}{l}\binom{M-K}{(n-1)-l}}{\binom{M-1}{n-1}} = 1$.
Substituting this expression into the one obtained for $\text{var}(X)$ it follows that

$$\begin{aligned}
\text{var}(X) &= \frac{nK}{M} \cdot \frac{(n-1)(K-1)}{M-1} + \frac{nK}{M} - \frac{n^2K^2}{M^2} \\
&= \frac{-n^2K^2(M-1) + Mn(n-1)K(K-1) + KnM(M-1)}{M^2(M-1)} \\
&= \frac{nK(M-K)(M-n)}{M^2(M-1)} \\
&= n\frac{K}{M}\left(1 - \frac{K}{M}\right)\frac{M-n}{M-1}.
\end{aligned}$$

and the result follows. □

4.2 Markov's inequality and Chebyshev's inequality

Markov's and Chebyshev's inequalities are two extremely important inequalities that will be used throughout this thesis. Markov's inequality provides an upper bound for the probability that a non negative random variable is greater than or equal to some positive constant. On the other hand Chebyshev's inequality gives an upper bound of the concentration of a random variable around it's expectation. Let us start by introducing Markov's inequality.

Theorem 4.17 (**Markov's inequality [1]**). *For every non-negative random variable X, and for all $k > 0$,*

$$P(X \geq k) \leq \frac{\mathbb{E}(X)}{k} .$$

Proof. Let I_k be the indicator variable for the event $X \geq k$ thus

$$I_k = \begin{cases} 1, \text{ if } X \geq k \\ 0, \text{ otherwise} . \end{cases}$$

Consider the random variable kI_k and note that

$$X \geq kI_k.$$

This is because if $X \geq k$ then $X \geq k \cdot 1 = kI_k$ and if $X < k$ then $X \geq k \cdot 0 = kI_k$. It follows that

$$\begin{aligned} \mathbb{E}(X) &\geq \mathbb{E}(kI_k) \qquad (\text{as} \quad X \geq I_k) \\ &= k\mathbb{E}(I_k) \\ &= k \cdot P(I_k = 1) \\ &= k \cdot P(X \geq k). \end{aligned}$$

The second equality follows from Theorem 3.21. □

Corollary 4.18. *If X is a non-negative random variable then for all $c \geq 1$,*

$$P(X \geq c \cdot \mathbb{E}(X)) \leq \frac{1}{c}.$$

Proof. Let $k = c\mathbb{E}(X)$. Using Markov's inequality we obtain the following.

$$\begin{aligned} P(X \geq c\mathbb{E}(X)) &\leq \frac{\mathbb{E}(X)}{\mathbb{E}(X)c} \\ &= \frac{1}{c} \end{aligned}$$

proving our corollary. □

Theorem 4.19 (**Chebyshev's inequality [1]**). *Let X be a random variable with finite expectation and finite variance* $\mathrm{var}(X)$. *Then for every $x > 0$ we have*

$$P(|X - \mathbb{E}(X)| \geq x) \leq \frac{\mathrm{var}(X)}{x^2}.$$

Proof. Consider the random variable $Y = (X - \mathbb{E}(X))^2$. Y is non-negative and $|X - \mathbb{E}(X)| \geq x$ if and only if $Y \geq x^2$. Using Markov's inequality we obtain that

$$\begin{aligned} P(|X - \mathbb{E}(X)| \geq x) &= P((X - \mathbb{E}(X))^2 \geq x^2) \\ &= P(Y \geq x^2) \\ &\leq \frac{\mathbb{E}(Y)}{x^2} \\ &= \frac{\text{var}(X)}{x^2}. \end{aligned}$$

The first inequality follows from Markov's inequality and the final equality from Definition 3.25. □

We now introduce Chebyshev's inequality and prove it using Markov's inequality.

Corollary 4.20. *Let $\sigma^2 = \text{var}(X)$ and let $\lambda > 0$ be a real number. Then $P(|X - \mathbb{E}(X)| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}$.*

Proof. Let $x = \lambda\sigma$. Using Chebyshev's inequality we obtain

$$\begin{aligned} P(|X - \mathbb{E}(X)| \geq \lambda\sigma) &\leq \frac{\text{var}(X)}{(\lambda\sigma)^2} \\ &= \frac{\sigma^2}{\lambda^2\sigma^2} \\ &= \frac{1}{\lambda^2}. \end{aligned}$$

Thus the result follows. □

In many cases we may want to bound the probability $P(X = 0)$. Markov's inequality and Chebyshev's inequality provide useful ways of doing this.

Theorem 4.21. *Let X be a non-negative integer random variable such that $\mathbb{E}(X) < 1$. Then*

$$P(X > 0) \leq \mathbb{E}(X).$$

Proof. Since $\mathbb{E}(X) < 1$, we use Markov's inequality to conclude that

$$P(X > 0) = P(X \geq 1) \leq \frac{\mathbb{E}(X)}{1} = \mathbb{E}(X),$$

and the result follows. □

The above theorem has important consequences. For example suppose X is dependent on some parameter n tending to infinity. If $\mathbb{E}(X) \to 0$ then $X = 0$ a.a.s (asymptotically almost surely) i.e with probability tending to 1 as n tends to infinity. On the other hand if $\mathbb{E}(X) \to \infty$ it does not necessarily mean that $X > 0$ a.a.s. We can deduce $X > 0$ a.a.s if further information is given.

Theorem 4.22. *Let X be a non negative random variable then*

$$P(X = 0) \leq \frac{\text{var}(X)}{\mathbb{E}(X)^2}.$$

Proof. It follows that

$$\begin{aligned} P(X = 0) &\leq P(|X - \mathbb{E}(X)| \geq \mathbb{E}(X)) \\ &= \frac{\text{var}(X)}{\mathbb{E}(X)^2}, \end{aligned}$$

where the second inequality follows from Chebyshev's inequality. □

The next Corollary will be one of the most important results in our thesis referred to as the second moment method.

Corollary 4.23. *If X is a non negative random variable and* $\text{var}(X) = o(\mathbb{E}(X)^2)$*, then $X > 0$ a.a.s.*

Proof. Using Theorem 4.22 it follows that

$$\begin{aligned} P(X = 0) &\leq \frac{\text{var}(X)}{\mathbb{E}(X)^2} \\ &= o(1). \end{aligned}$$

Hence

$$\begin{aligned} P(X > 0) &= 1 - P(X = 0) \\ &= 1 - o(1), \end{aligned}$$

thus $X > 0$ a.a.s. □

We can prove a much stronger claim than the one seen in Corollary 4.23.

Theorem 4.24. *If X is a non negative random variable and $\text{var}(X) = o(\mathbb{E}(X)^2)$, then a.a.s $X \approx \mathbb{E}(X)$.*

Proof. Let $\epsilon > 0$ be arbitrarily small but fixed. Using Chebyshevs inequality

$$\begin{aligned} P(|X - \mathbb{E}(X)| \geq \epsilon \mathbb{E}(X)) &\leq \frac{\text{var}(X)}{(\epsilon \mathbb{E}(X))^2} \\ &= \frac{\text{var}(X)}{\epsilon^2 \mathbb{E}(X)^2} \\ &= \frac{1}{\epsilon^2} o(1) \\ &= o(1). \end{aligned}$$

We conclude that a.a.s

$$(1 - \epsilon)\mathbb{E}(X) \leq X \leq (1 + \epsilon)\mathbb{E}(X)$$

and thus $X \approx \mathbb{E}(X)$ a.a.s. □

Theorem 4.25. *Let A_i be a set of events where $1 \leq i \leq n$. Let X_i be the indicator random variable for the event A_i and let $X = X_1 + X_2 + \ldots + X_m$ be a non negative random variable. Then*

$$\text{var}(X) \leq \mathbb{E}(X) + \sum_{i \neq j} \text{cov}(X_i, X_j).$$

Proof. Suppose $P(A_i) = p_i$, then given the indicator random variable X_i we find using Theorem 4.3 that

$$\mathbb{E}(X_i) = P(A_i) = p_i$$

and

$$\text{var}(X_i) = p_i(1 - p_i).$$

Furthermore as $(1 - p_i) \leq 1$ it follows that

$$\text{var}(X_i) = p_i(1 - p_i) \leq p_i = \mathbb{E}(X_i).$$

Finally, from Theorem 3.39

$$\begin{aligned} \text{var}(X) &= \sum_{i=1}^{m} \text{var}(X_i) + \sum_{i \neq j} \text{cov}(X_i, X_j) \\ &\leq \sum_{i=1}^{n} \mathbb{E}(X_i) + \sum_{i \neq j} \text{cov}(X_i, X_j) \\ &= \mathbb{E}(X) + \sum_{i \neq j} \text{cov}(X_i, X_j) \end{aligned}$$

with the final equality following from the linearity of expectation. □

Definition 4.26. Let A_i be a set of events where $1 \leq i \leq n$. Let X_i be the indicator random variable for the event A_i and let $X = X_1 + X_2 + \ldots + X_m$. For indices i, j write $i \sim j$ if $i \neq j$ and the events A_i and A_j are dependent. Denote

$$\Delta = \sum_{i \sim j} P(A_i \cap A_j).$$

Theorem 4.27. *Let X_i be a non-negative random variables where $1 \leq i \leq m$ and let the random variable $X = X_1 + X_2 + \ldots + X_m$. If $\mathbb{E}(X) \to \infty$ and $\Delta = o(\mathbb{E}(X)^2)$ then $X > 0$ a.a.s, and furthermore $X \approx \mathbb{E}(X)$ a.a.s.*

Proof. Firstly note that if $i \sim j$, then

$$\begin{aligned} \text{cov}(X_i, X_j) &= \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) \\ &\leq \mathbb{E}(X_i X_j) \\ &= P(A_i \cap A_j). \end{aligned}$$

Furthermore if $i \neq j$ and $i \not\sim j$ then

$$\text{cov}(X_i, X_j) = 0.$$

Hence,

$$\begin{aligned}\operatorname{var}(X) &\leq \mathbb{E}(X) + \sum_{i \neq j} \operatorname{cov}(X_i X_j) \\ &\leq \mathbb{E}(X) + \sum_{i \sim j} \mathbb{E}(X_i X_j) \\ &= \mathbb{E}(X) + \sum_{i \sim j} P(A_i \cap A_j) \\ &= \mathbb{E}(X) + \Delta \\ &= o(\mathbb{E}(X)^2),\end{aligned}$$

with the first inequality following from Theorem 4.25 and the final equality follows from the fact that $\mathbb{E}(X) \to \infty$ and $\Delta = o(\mathbb{E}(X)^2)$. Finally using Theorem 4.24 it follows that $X \approx \mathbb{E}(X)$ a.a.s.

□

Definition 4.28. Let us say the indicator random variables $X_1, \ldots, X_m$ for events $A_1, \ldots, A_m$ are symmetric if for every $i \neq j$ there is an automorphism of the underlying probability space that sends event A_i to event A_j.

Examples will appear in our final chapter. In the case that $X_1, \ldots, X_m$ are symmetric denote

$$\Delta = \sum_{i \sim j} P(A_i \cap A_j) = \sum_i P(A_i) \sum_{j \sim i} P(A_j | A_i)$$

note the inner summation is independent of i. Set $\Delta^* = \sum_{j \sim i} P(A_i | A_j)$ where i is any fixed index. Thus

$$\Delta = \sum_i P(A_i) \Delta^* = \mathbb{E}(X) \Delta^*.$$

The following is a Corollary of Theorem 4.27.

Corollary 4.29. *Let $X_1, \ldots, X_m$ be symmetric indicator random variables for the events $A_1, \ldots, A_m$. Let $X = \sum_{i=1}^m X_i$. If $\mathbb{E}(X) \to \infty$ and $\Delta^* = o(\mathbb{E}(X))$ then $X > 0$ a.a.s, furthermore $X \approx \mathbb{E}(X)$.*

Proof. If $\Delta^* = o(\mathbb{E}(X))$ then $\Delta = o(\mathbb{E}(X)^2)$ and the result follows from Theorem 4.27. □

4.3 Moment generating functions

It is possible to obtain substantially stronger bounds than Markov and Chebyshev when we have more information about the random variable X. Given the random variable X consider the random variable $Y = e^{sX}$ where $s \geq 0$ is a parameter. If we now apply Markov's inequality to the non-negative random variable Y we find that

$$P(X \geq t) = P(e^{sX} \geq e^{st}) \leq \frac{\mathbb{E}(e^{sX})}{e^{st}}.$$

Provided we can show $\mathbb{E}(e^{sX})$ is not large, the bound obtained has a denominator which grows exponentially which is significantly better than the Chebyshev's bound which grows quadratically.

Definition 4.30. For any non-negative integer k define the k-th moment of the random variable X, to be the function

$$M_k = \mathbb{E}(X^k).$$

Definition 4.31. The moment-generating function of X in the indeterminate s is the function

$$M_X(s) = \mathbb{E}(e^{sX}).$$

We define $M_X^k(s)$ to be the k-th derivative of the moment-generating function evaluated at s. The following theorem states that one can compute the k-th moment of X by evaluating the k-th derivative of M_X evaluated at $s = 0$.

Theorem 4.32. *Let X be a random variable. Then*

$$\mathbb{E}(X^k) = M_X^k(0).$$

Proof. We prove the following stronger claim that

$$M_X^k(s) = \mathbb{E}(X^k e^{sX}).$$

We proceed by induction on k. The base case $k = 0$ is

$$\begin{aligned} M_X^0(s) &= M_X(s) \\ &= \mathbb{E}(e^{sX}) \\ &= \mathbb{E}(X^0 e^{sX}). \end{aligned}$$

Now for the inductive step suppose the claim holds for all $i \leq k$.

$$\begin{aligned} M_X^{k+1}(s) &= \frac{d}{ds} M_X^k(s) \\ &= \frac{d}{ds} \mathbb{E}(X^k e^{sX}) \\ &= \mathbb{E}(\frac{d}{ds} X^k e^{sX}) \\ &= \mathbb{E}(X^{k+1} e^{sX}). \end{aligned}$$

Hence the claim holds for $k + 1$. Finally evaluating at $s = 0$ we have

$$M_X^k(0) = \mathbb{E}(X^k e^{0X}) = \mathbb{E}(X^k).$$

and the result holds. □

Theorem 4.33. *Let $X_1, X_2, \ldots, X_n$ be independent random variables and let $X = X_1 + \ldots + X_n$. Then*

$$M_X(s) = \prod_{i=1}^{n} M_{X_i}(s).$$

Proof.

$$\begin{aligned} M_X(s) &= \mathbb{E}\left(e^{s(X_1+\ldots+X_n)}\right) \\ &= \mathbb{E}\left(\prod_{i=1}^{n} e^{s(X_i)}\right) \\ &= \prod_{i=1}^{n} \mathbb{E}(e^{s(X_i)}) \\ &= \prod_{i=1}^{n} M_{X_i}(s). \end{aligned}$$

The third equality follows from Theorem 3.23. □

Theorem 4.34. *Let X be any random variable. Then, for any $t > 0$ we have*

1. $P(X \geq t) \leq \min_{s>0} \frac{M_X(s)}{e^{st}}$
2. $P(X \leq t) \leq \min_{s<0} \frac{M_X(s)}{e^{st}}$

Proof. Starting with 1), for any $s > 0$ we have,

$$\begin{aligned} P(X \geq t) &= P(e^{sX} \geq e^{st}) \\ &\leq \frac{\mathbb{E}(e^{sX})}{e^{st}} \\ &= \frac{M_X(s)}{e^{st}}. \end{aligned}$$

The first inequality follows from Markov's inequality.

Next we prove part 2) for any $s < 0$,

$$\begin{aligned} P(X \leq t) &= P(e^{sX} \geq e^{st}) \\ &\leq \frac{\mathbb{E}(e^{sX})}{e^{st}} \\ &= \frac{M_X(s)}{e^{st}}. \end{aligned}$$

The first inequality follows from Markov's inequality. Since the bound holds for every s in some range, it holds for the minimum over this range. □

Corollary 4.35. *Let $X = \sum_{i=1}^{n} X_i$, where X_i are independent. Then for any $t > 0$ we have*

$$P(X \geq t) \leq \min_{s \geq 0} \frac{\prod_{i=1}^{n} M_{X_i}(s)}{e^{st}}$$

and

$$P(X \leq t) \leq \min_{s \leq 0} \frac{\prod_{i=1}^{n} M_{X_i}(s)}{e^{st}}.$$

Proof. We combine Corollary 4.33 and Theorem 4.34 to obtain

$$\begin{aligned} P(X \geq t) &\leq \min_{s \geq 0} \frac{M_X(s)}{e^{st}} \\ &= \min_{s \geq 0} \frac{\prod_{i=1}^{n} M_{X_i}(s)}{e^{st}}. \end{aligned}$$

Similarly

$$\begin{aligned} P(X \leq t) &\leq \min_{s \leq 0} \frac{M_X(s)}{e^{st}} \\ &= \min_{s \leq 0} \frac{\prod_{i=1}^{n} M_{X_i}(s)}{e^{st}}. \end{aligned}$$

□

4.4 Bounds with moment Generating functions: Chernoff Bounds

We are now interested in using the inequalities obtained using moment generating functions to obtain bounds given specific types of random variables. We consider random variables X where $X = \sum_{i=1}^{n} X_i$ and the X_i are independent Bernoulli random variables.

Lemma 4.36. *Let X_i be Bernoulli random variable with*

$$P(X_i = 1) = p \text{ and } P(X_i = 0) = 1 - p.$$

Then

$$M_{X_i}(s) \leq e^{p(e^s - 1)}.$$

Proof. We use the inequality $1 + x \leq e^x$ which holds for all x. Now

$$\begin{aligned} M_{X_i}(s) &= \mathbb{E}(e^{sX_i}) \\ &= (1-p)e^{s\cdot 0} + pe^{s\cdot 1} \\ &= 1 + p(e^s - 1) \\ &\leq e^{p(e^s-1)} \end{aligned}$$

and the result follows □

Lemma 4.37. *Let $X = \sum_{i=1}^n X_i$ where each X_i is an independent Bernoulli random variable with $P(X_i = 1) = p_i$. Then*

$$M_X(s) \leq e^{(e^s-1)\mathbb{E}(X)}.$$

Proof.

$$\begin{aligned} M_X(s) &= \prod_{i=1}^n M_{X_i}(s) \\ &\leq \prod_{i=1}^n e^{(e^s-1)p_i} \\ &= e^{(e^s-1)\sum_{i=1}^n p_i} \\ &= e^{(e^s-1)\mathbb{E}(X)}. \end{aligned}$$

The first equality follows from Theorem 4.33 and the first inequality follows from Lemma 4.36. □

Theorem 4.38 (Chernoff-uppertail)**.** *Let $X = \sum_{i=1}^n X_i$, where the X_i are independent Bernoulli random variables with $P(X_i = 1) = p_i$. Then for any $\delta > 0$ we have*

$$P(X \geq (1+\delta)\mathbb{E}(X)) \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{\mathbb{E}(X)}.$$

Proof. Set $t = (1+\delta)\mathbb{E}(X)$ and $s = \ln(1+\delta) > 0$. Then using Theorem 4.34 and Lemma 4.37 we obtain

$$\begin{aligned} P(X \geq (1+\delta)\mathbb{E}(X)) &= P(X \geq t) \\ &\leq \min_{s\geq 0} \frac{e^{(e^s-1)\mathbb{E}(X)}}{e^{(s(1+\delta))\mathbb{E}(X)}} \\ &\leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{\mathbb{E}(X)}. \end{aligned}$$

The first inequality follows from Theorem 4.34 and Lemma 4.37. The second inequality follows from substituting $s = \ln(1+\delta)$ into the first inequality. □

Using Theorem 4.38 we obtain the following slightly weaker but easier to apply bound.

Corollary 4.39 (Chernoff-uppertail)**.** *For* $0 < \delta \leq 1$

$$P(X \geq (1+\delta)\mathbb{E}(X)) \leq e^{-\mathbb{E}(X)\frac{\delta^2}{3}}.$$

Proof. It suffices to prove that for every $0 < \delta \leq 1$

$$\frac{e^\delta}{(1+\delta)^{1+\delta}} \leq e^{-\frac{\delta^2}{3}}.$$

Taking the natural logarithm of both sides of the inequality, it suffices to show that

$$f(\delta) := \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{3} \leq 0.$$

Now it follows that

$$\begin{aligned} f'(\delta) &= \delta' - (1+\delta)'\ln(1+\delta) - (1+\delta)(\ln(1+\delta))' + (\left(\frac{\delta^2}{3}\right)' \\ &= 1 - \ln(1+\delta) - (1+\delta)\cdot\frac{1}{1+\delta} + \frac{2\delta}{3} \\ &= -\ln(1+\delta) + \frac{2\delta}{3}. \end{aligned}$$

Furthermore

$$\begin{aligned} f''(\delta) &= -(\ln(1+\delta))' + \left(\frac{2\delta}{3}\right)' \\ &= -\frac{1}{1+\delta} + \frac{2}{3}. \end{aligned}$$

Now we observe that

$$f''(\delta) < 0 \text{ for } 0 < \delta \leq \frac{1}{2}$$

and

$$f''(\delta) \geq 0 \text{ for } \frac{1}{2} \leq \delta < 1.$$

So $f'(\delta)$ decreases in the interval $[0, \frac{1}{2})$ and then increases in the interval $[1/2, 1]$. Finally as $f'(0) = 0$ and $f'(1) < 0$ it follows that $f'(\delta) < 0$ for all $\delta \in [0, 1]$. As $f(0) = 0$ we conclude $f(\delta) \leq 0$ for all $\delta \in [0, 1]$. □

Theorem 4.40 (Chernoff-lower tail)**.** *Let $X = \sum_{i=1}^{n} X_i$, where the X_i are independent Bernoulli random variables with $P(X_i = 1) = p_i$. Then for any $0 < \delta < 1$ we have*

$$P(X \geq (1-\delta)\mathbb{E}(X)) \leq \left(\frac{e^{\delta}}{(1-\delta)^{1-\delta}}\right)^{\mathbb{E}(X)}.$$

Proof. Set $t = (1-\delta)\mathbb{E}(X)$ and $s = \ln(1-\delta) < 0$. Then using Theorem 4.34 and Lemma 4.37 we obtain

$$\begin{aligned} P(X \leq (1-\delta)\mathbb{E}(X)) &\leq \min_{s \leq 0} \frac{e^{(e^s-1)\mathbb{E}(X)}}{e^{(s(1+\delta))\mathbb{E}(X)}} \\ &\leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mathbb{E}(X)}. \end{aligned}$$

The first inequality follows from Theorem 4.34 and Lemma 4.37. The second inequality follows from substituting $s = \ln(1-\delta)$ into the first inequality. □

Similarly to Corollary 4.39 we obtain a slightly weaker but easier to apply bound for the lower tail.

Corollary 4.41 (Chernoff-lower tail)**.** *For* $0 < \delta < 1$

$$P(X \leq (1-\delta)\mathbb{E}(X)) \leq e^{-\mathbb{E}(X)\frac{\delta^2}{2}}$$

Proof. It suffices to prove that for every $0 < \delta < 1$

$$\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \leq e^{\frac{\delta^2}{2}}.$$

Taking the natural logarithm of both of sides of the inequality, it suffices to show that

$$f(\delta) := -\delta - (1-\delta)\ln(1-\delta) + \frac{\delta^2}{2} \leq 0.$$

Now it follows that

$$\begin{aligned} f'(\delta) &= -\delta' - (1-\delta)'\ln(1-\delta) - (1-\delta)(\ln(1-\delta))' + \left(\frac{\delta^2}{2}\right)' \\ &= -1 + \ln(1-\delta) - (1-\delta)\cdot\frac{1}{1-\delta} + \delta \\ &= -2 + \ln(1-\delta) + \delta. \end{aligned}$$

Furthermore,

$$\begin{aligned} f''(\delta) &= -(2)' + (\ln(1-\delta))' + \delta' \\ &= 1 - \frac{1}{(1-\delta)}. \end{aligned}$$

We obtain that $f''(\delta) < 0$ for all $0 < \delta < 1$ and so $f'(\delta)$ is decreasing over the interval $[0, 1]$. As $f(0) = 0$ and $f'(0) = -2$ because $f'(\delta)$ decreases over the interval $[0, 1]$ it follows that $f(\delta) \leq 0$ for all $0 < \delta < 1$. □

Corollary 4.42. *If* $X \sim \text{Bin}(n, p)$*, then for any* $\delta > 0$ *we have*

$$P(X \geq (1+\delta)np) \leq e^{\frac{-\delta^2 np}{3}}.$$

Proof. By noting $\mathbb{E}(X) = np$ from Theorem 4.6, the result follows from Corollary 4.39.

□

Corollary 4.43. *If $X \sim \text{Bin}(n, p)$ then for any $0 < \delta < 1$ we have*

$$P(X < (1 - \delta)np) \leq e^{\frac{-\delta^2 np}{2}}.$$

Proof. The proof is similar to that of Corollary 4.42. □

Chapter 5

Graph theory

The Königsberg bridge problem is a famous problem in mathematics which asks if the seven bridges of Königsberg can all be crossed without crossing the same bridge more than once, with the trip ending in the same place it started. The problem was solved by Euler in 1736 and his solution represented the beginning of graph theory. Informally a graph is a collection of dots known as vertices and lines known as edges that connect the vertices. After representing the map of Königsberg and it's seven bridges by a collection of vertices and edges, careful study of the graph by Euler lead him to solve the problem with a negative solution. Since then graph theory has become a major field of study in both mathematics and theoretical computer science. It's framework is responsible for the solution of many problems in areas such as optimization, enumeration and existence. In addition many famous problems have been solved using graph theoretic techniques, such as the knights-tour problem and the four colour problem, the latter of which we will briefly discuss in this chapter.

5.1 Introduction to graph theory

We start by formally introducing many of the graph theoretic terms we will use throughout this chapter. Let us start by first formally introducing the concept of a graph.

Definition 5.1. A graph $G = (V, E)$ is a pair of sets with $E \subseteq [V]^2$. Thus elements of E are two element subsets of V. A single vertex is known as a trivial graph.

As mentioned above the elements of V are known as vertices and the elements of E as edges. A graph with vertex set V is said to be a graph on V. Given a graph $G = (V, E)$ we call the number of vertices in G its order and often denote this as $|G|$ instead of $|V|$. The order of the graphs we will encounter in this work will be finite.

Two vertices v_1, v_2 are said to be adjacent if $\{v_1, v_2\} \in E$ and are independent otherwise. We shall for simplicity refer to the edge $\{v_1, v_2\}$ as v_1v_2. Two edges e_1, e_2 are incident if they share a common vertex. We shall also say a vertex v is incident to an edge e if $v \in e$.

Definition 5.2. Given a graph $G = (V, E)$, an independent set is a set of vertices no two of which are adjacent.

Definition 5.3. Given a graph $G = (V, E)$, a maximum independent set is an independent set of largest possible size in G. It's size is called the independence number of G and is denoted by $\alpha(G)$.

In this chapter and the remaining chapters any reference to a graph will be a reference to a labeled graph. A labeled graph is a graph where the vertices have unique labels and in this thesis unless otherwise stated a labeled graph on n vertices will have vertex set $V = [n]$.

The first type of graph we introduce is a complete graph. A complete graph is a graph which contains all possible edges. A complete graph on n vertices is denoted by K_n. It follows that a complete graph on n vertices contains exactly $\binom{n}{2}$ edges, since each 2 element subset of $[n]$ is an edge of K_n.

Example 5.4. The complete graph on 4 vertices K_4 has edge set

$$E = \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}.$$

Definition 5.5. Given a graph $G = (V, E)$ a clique is a set of vertices such that every two vertices in the set are adjacent.

Definition 5.6. Given a graph $G = (V, E)$ a maximum clique is a clique of largest possible size in G. It's size is called the clique number of G and is denoted by $\omega(G)$.

Definition 5.7. Two graphs $G = (V, E)$ and $G' = (V', E')$ are isomorphic denoted $G \cong G'$, if there exists a bijection $\eta : V \to V'$ such that $xy \in E$ if and only if $\eta(x)\eta(y) \in E'$ for all $x, y \in V$.

In the remainder of this thesis we will often be interested in classes of graphs which have the property of containing another graph as a substructure. Let us now define the notion of a subgraph.

Definition 5.8. Given two graphs $G = (V, E)$ and $H = (V', E')$ we say that H is a subgraph of G denoted $H \subseteq G$, if $V' \subseteq V$ and $E' \subseteq E$.

Definition 5.9. Given a graph $G = (V, E)$ and subgraph $G' = (V', E') \subseteq G$. If G' contains precisely those edges in G with both ends in V' then we say that G' is an induced subgraph denoted $G' = G[V']$.

Given a vertex $v \in V$ in a graph $G = (V, E)$ it is of particular interest to know how many other vertices v is adjacent to. This leads us to the idea of vertex degree and the set of neighbours of a vertex.

Definition 5.10. Given a graph $G = (V, E)$ and a vertex $v \in V$. The set

$$N(v) = \{x \in V : xv \in E\}$$

is called the set of neighbours of v.

Definition 5.11. Given a graph $G = (V, E)$ and a vertex $v \in V$, the degree of the vertex $d(v)$ is the number of neighbours of v, $d(v) = |N(v)|$. We denote $\delta(G) = \min\{d(v)|v \in V\}$ and $\Delta(G) = \max\{d(v)|v \in V\}$ as the minimum and maximum degree of G respectively. If $d(v) = 0$ we say that v is isolated.

Definition 5.12. We define the average degree of a graph $G = (V, E)$ as $\frac{1}{|V|}\sum_{v\in V} d(v)$ and denote it by $d(G)$.

We can workout the average degree of a graph given the number of edges and vica-versa using the following result.

Proposition 5.13. *Given a graph $G = (V, E)$*

$$|E| = \frac{1}{2}\sum_{v\in V} d(v) = \frac{1}{2}d(G)|V|.$$

Proof. Each edge $v_1v_2 \in E$ is counted twice in $\sum_{v\in V} d(v)$, once in $d(v_1)$ and once in $d(v_2)$. Thus

$$2|E| = \sum_{v\in V} d(v) = d(G)|V|,$$

and the result follows. □

Proposition 5.14. *The number of vertices of odd degree in a graph is always even.*

Proof. As $\sum_{v\in V} d(v) = 2|E|$ the sum must is even. Hence the number of vertices with odd degree must be even. □

5.1.1 Paths and Cycles

The notion of a path and cycle are fundamental to graph theory. We start by formally introducing both concepts.

Definition 5.15. A path of length n denoted P_{n+1} is a graph with vertex set $V = \{v_0, v_1, \ldots, v_n\}$ and edge set $E = \{v_0v_1, v_1v_2, \ldots, v_{n-1}v_n\}$. We will denote P_{n+1} as

$$P_{n+1} = v_0v_1 \ldots v_n.$$

Definition 5.16. A cyle of length $n+1$ is a graph with vertex set $V = \{v_0, v_1, \ldots, v_n\}$ and edge set $E = \{v_0v_1, v_1v_2, \ldots, v_{n-1}v_n, v_nv_0\}$. We will denote C_{n+1} as

$$C_{n+1} = v_0v_1 \ldots v_nv_1.$$

Given a path $P = x_1 \ldots x_n$ we shall refer to the subpath between points x_i and x_j, $i \neq j$ on P as x_iPx_j.

It is often of interest to know the size of the smallest cycle contained within a graph.

Definition 5.17. The length of the smallest cycle in a graph $G = (V, E)$ is known as it's girth denoted $g(G)$. If a graph contains no cycle we say it has a girth of ∞.

The first result we prove in this section, gives us a sufficient condition involving the minimum degree of a graph for the existence of a cycle. This result will prove useful in later sections.

Proposition 5.18. *Every graph G with $\delta(G) \geq 2$ has a path of length at least $\delta(G)$ and a cycle of length at least $\delta(G) + 1$.*

Proof. Let $P = x_0 \ldots x_k$ be a longest path in G, then all neighbors of x_k must lie on P otherwise we could obtain a longer path in G. It then follows that and so $\delta(G) \leq d(x_k) \leq k$. Thus there exists a path of length at least $\delta(G)$.
Suppose $i < k$ is minimal with $x_ix_k \in E(G)$. Then $x_iCx_kx_i$ is a cycle of length at least $\delta(G) + 1$. □

5.1.2 Connectivity

The idea of a connected graph is an intuitive one. We say that a graph is connected if for any pair of vertices in $G = (V, E)$ there exists a path between them. Connectivity will be studied in greater detail in the latter chapters when we introduce the topics of Random Graphs. Let us start by defining some key concepts in the study of connectivity.

Definition 5.19. Given a $G = (V, E)$ an inclusion maximal connected subgraph of G is called a component of G.

Example 5.20. The graph $G = ([5], \{\{1, 2\}, \{3, 4\}, \{4, 5\}\})$ has 2 components. Namely the edge $\{1, 2\}$ and the induced subgraph on vertex set $\{3, 4, 5\}$ with edges $\{3, 4\}$ and $\{4, 5\}$.

Proposition 5.21. *The vertices of a connected graph G can always be enumerated as $v_1, \ldots, v_n$ so that $G_i = G[v_1, v_2, \ldots, v_i]$ is connected for every i.*

Proof. We proceed by induction on $i < |G|$, first note that a single vertex is viewed as being a connected graph. Assume inductively that $v_1, \ldots v_i$ have been chosen for some $i < |G|$ and arbitrarily pick a vertex $v \in V(G) \setminus V(G_i)$. As G is connected, there is a path P from v to v_1. Choose $v_{i+1} \in V(G) \setminus V(G_i)$ to be some vertex on P whose neighbour is in $V(G_i)$. Hence $G[v_1, v_2, \ldots, v_{i+1}]$ is connected.

□

Given a connected graph it is useful to obtain a measure of exactly how connected it is. For example can we disconnect it by removing a vertex? Or will it still remain connected? The definition of a k-connected graph follows from trying to obtain a measure of a graphs connectedness.

Definition 5.22. A graph $G = (V, E)$ is called k-connected for $k \in \mathbb{N}$ if $|G| \geq k$ and removing any set $k - 1$ vertices does not disconnect G.

Definition 5.23. The greatest integer k for which $G = (V, E)$ is k connected is called the connectivity number of G denoted $\kappa(G)$.

We now introduce a similar notion involving the removal of edges.

Definition 5.24. A graph $G = (V, E)$ is called l-edge connected if for every set of $F \subseteq E$ edges with $|F| < l$ the graph $G = (V, E \setminus F)$ is connected.

Definition 5.25. The greatest integer l such that $G = (V, E)$ is l-edge connected is called the edge-connectivity number of G denoted $\lambda(G)$ (if G is disconnected then $\lambda(G) = 0$).

Definition 5.26. Given a set of edges F of a graph $G = (V, E)$. If the graph $G' = (V, E \setminus F)$ is disconnected, we say that F is an edge separator. F is a minimal edge separator if $|F| = \lambda(G)$.

The connectivity number and edge-connectivity number of a graph are related.

Theorem 5.27. *For every non trivial graph $G = (V, E)$ it holds that*

$$\kappa(G) \leq \lambda(G) \leq \delta(G).$$

Proof. Let $G = (V, E)$ be a graph on n vertices and x a vertex of minimum degree. Then removing the edges incident to the vertex x form an edge separator thus $\lambda(G) \leq \delta(G)$.
The vertex connectivity of any graph on n vertices can be bounded above by $\kappa(K_n) = n - 1$, thus $\kappa(G) \leq n - 1$. Consider a minimal edge separator in $G = (V, E)$ that partitions the vertex set into two sets A and $B = V \setminus A$. If all edges between these two sets exist then $\lambda(G) = |A||B| \geq n - 1$. If not there exists vertices $x \in A$ and $y \in B$ with no edge between them. If we remove from G, the neighbours of x in B and the set of vertices in A which have neighbours in B-then we seperate x and y and thus the resulting graph is disconnected. Suppose we remove r vertices in total, it follows that $\kappa(G) \leq r$ while the number of edges between A and B is atleast r. Hence the result follows. □

5.1.3 Trees and Forests

We have introduced paths and cycles, the next specific type of graph we look at is a tree.

Definition 5.28. A graph not containing any cycles is called a forest. A connected forest is called a tree.

The term tree was coined by British mathematician Arthur Cayley in 1857. Cayley's formula tells us that there are n^{n-2} trees on n labeled vertices. While we do not prove this in this thesis we will refer to it at a later stage.

Definition 5.29. The vertices of a tree with degree 1 are called its leaves.

Proposition 5.30. *Every tree contains a leaf.*

Proof. It follows from Proposition 5.18 that any graph satisfying $\delta(G) \geq 2$ contains a cycle. Thus any tree T satisfies $\delta(T) = 1$ since a tree is connected a graph. □

Theorem 5.31. *The following assertions are equivalent for a graph* $T = (V, E)$.

1. T is a tree.
2. Any two vertices of T are linked by a unique path in T.
3. T is minimally connected, i.e T is connected but for any edge $e \in E$, $T' = (V, E \setminus \{e\})$ is disconnected.
4. T is maximally acyclic, i.e T contains no cycle but $T'' = (V, E \cup xy)$ does for any two non-adjacent vertices $x, y \in V$.

Proof. $(1 \to 2)$
We prove that if condition 2 does not hold then condition 1 cannot hold either. Take a graph $G = (V, E)$ which does not satisfy condition 2). Then either V contains a pair of vertices x, y with no path between them, or with at least two disjoint paths between them. If there exists no path between x and y then G is not connected and thus is not a tree. Suppose instead there exists two disjoint paths P_1 and P_2 between vertices x and y. Let z be the first vertex where the two paths intersect moving x to y (note y maybe such a point). Then $C = xP_1zP_2x$ forms a cycle in G and so it is not a tree.
$(2 \to 3)$
We prove that if every pair of vertices in a graph $T = (V, E)$ are linked by a unique path then T is minimally connected. First of all since there is a path connecting every pair of vertices in T it follows that T is connected.

For each edge $e = xy$ of T, the unique path from x to y is the edge e itself. Thus removing any edge of T will result in it's end points being disconnected. Hence $T' = (V, E \setminus \{e\})$ is disconnected.
$(3 \to 4)$
If T is minimally connected then it does not contain a cycle. Otherwise we can remove an edge from any cycle and the resulting graph will still be connected contradicting that it is minimally connected. Furthermore since T is connected there is a path between any two vertices x and y. Adding the edge $e = xy$ will thus form a cycle and so T is maximally acyclic.
$(4 \to 1)$
If T is acyclic then it must be a forest. Furthermore if T is maximally acyclic it must be connected as otherwise we can add an edge between two components and avoid adding a cycle. Hence T is a tree. □

An interesting feature of all tree's on n vertices is that they have the same number of edges, let us take a look at this in more detail.

Proposition 5.32. *A connected graph with n vertices is a tree if and only if it has $n - 1$ edges.*

Proof. We first prove inductively that if a graph on n vertices is connected and has $n - 1$ edges then it is a tree. Clearly the case $n = 2$ holds. Suppose that the claim holds for all graphs on $i < n$ vertices and consider any graph $G = (V, E)$ on n vertices. It follows from Proposition 5.13 that G must contain a leaf otherwise $\delta(G) \geq 2$ and the number of edges in G would be at least n contradicting that G has $n - 1$ edges. Let $x \in V$ with $d(x) = 1$, and consider the graph G' obtained by removing x from G. The graph G' is a graph on $n-1$ vertices with $n-2$ edges and thus is a tree by our inductive hypothesis. It follows that G is also a tree since it is obtained by adding a leaf to G'.

Next we use induction again, but this time to prove that if T is a tree on n vertices then it has $n - 1$ edges. The case $n = 2$ is trivial. Suppose the the claim holds for all graphs on $i < n$ vertices and consider any tree $T = (V, E)$ on n vertices. Let $x \in V$ such that $d(x) = 1$ and consider the tree T' obtained by removing the leaf x. Since T' is tree on $n - 1$ vertices it follows from our inductive hypothesis that T has $n - 2$ edges. Since T is obtained from T' by adding a leaf, we can conclude that T has $n - 1$ edges. □

Definition 5.33. Given a graph $G = (V, E)$ a spanning tree T is a subgraph which is a tree and contains each vertex of V.

Proposition 5.34. *Every connected graph $G = (V, E)$ contains a spanning tree T.*

Proof. If G contains no cycles then G is a tree and therefore is by definition a spanning tree. If G contains a cycle C then obtain a new graph G' by removing an edge of C, note G' is still connected. If G' does not contain a cycle then G' is a spanning tree for G. If G' contains a cycle C' obtain a new graph G'' by removing an edge of C', again G'' is connected. As G can only contain a finite number of cycles this process must stop with a graph T which does not contain a cycle and is connected. Hence T is a spanning tree. □

5.1.4 Bipartite graphs

Definition 5.35. Let $r \geq 2$ be an integer, a graph $G = (V, E)$ is called r-partite if V admits a partition into r classes such that vertices in the same class are not adjacent. For the case $r = 2$ we say that the graph is bipartite.

Proposition 5.36. *A graph is bipartite if and only if it contains no odd cycle.*

Proof. We start by proving that if $G = (V, E)$ contains no odd cycles then it's bipartite. If all components of a disconnected graph are bipartite then the graph itself is bipartite, hence we can assume wlog that $G = (V, E)$ is a connected. Consider a spanning tree $T = (V', E')$ of G and pick a vertex $r \in V'$. We now partition $V(G)$ into two vertex classes A and B as follows. For all $y \in V$ if the shortest path from r to y has odd length, we put y into class A and otherwise into class B (note r has length 0 from itself and so goes into class B). If for each edge $e = ab \in E$ of G, a and b are in different classes then our graph G is bipartite. Suppose the edge $e = xy$ is an edge of T, then x and y are in different vertex classes. If $e' = x'y'$ is an edge of G but not an edge of T, then we have a cycle $C = x'Ty'x'$ in G. If x' and y' are both in the same class then, the cycle C has odd length. However this contradicts that G contains no odd cycles and so x' and y' are in different

classes and so for each edge $e = ab$ it holds that a and b are in different classes and it follows that $G = (V, E)$ is bipartite.

Next we show that if $G = (V, E)$ is a graph which contains an odd cycle it cannot be bipartite. Let $C = x_1x_2 \dots x_{2k+1}x_1$ be an odd cycle in G. Each vertex of C must be assigned a class A or B with adjacent vertices being assigned to different classes. It follows by the pigeon hole principle that at least $k + 1$ of these vertices must be assigned the same colour, however then two adjacent vertices must be assigned the same colour and so G is not a bipartite graph. □

5.2 Chromatic Number

Example 5.37. What is the smallest number of colours required to colour the countries of a map to ensure adjacent countries recieve different colours?.

Given a map we construct a graph $G = (V, E)$ as follows. Let the countries be our vertices and join two vertices by an edge if their corresponding countries are adjacent, the problem then reduces to finding the smallest number of colours needed to colour our graph G. It turns out that our graph G is a planar graph (a graph which can be drawn without any edges crossing) and the problem is equivalent to the famous 4 colour problem in graph theory. The problem states that every planar graph can be coloured with 4 colours such that adjacent vertices receive different colours.

Example 5.38. Suppose we have n committees and if two committees are opposed they cannot operate from the same town, what is the minimum number of towns we need to house all the committees?

While the first two examples seem unrelated, we can attempt to solve them using the same graph theoretic principles. In the second example we construct a graph $G' = (V', E')$ as follows, let our committees be our vertices and we join two vertices by an edge if their corresponding committees are opposed. Finding the smallest number of towns to house the committees is equivalent to finding the least number of colours required to colour the

vertices of G' such that adjacent vertices receive different colours (where the colours represent the towns).

Both of the problems in the examples above, require us to colour vertices in such a way that adjacent vertices receive different colours. Of course we can always use n colours to colour n vertices of a graph, but we are required to use the least number of colours possible. Solutions to such problems are highly sort after in areas such as scheduling theory, and as seen above we can use graph theoretic principles to solve them. This leads us to introduce the idea of vertex colouring more formally.

Definition 5.39. A k vertex colouring (or k-colouring) of a graph $G = (V, E)$ is a map $c : V \to \{1, 2, \ldots, k\}$ such that $c(v) \neq c(w)$ if v and w are adjacent.

Definition 5.40. Given a graph $G = (V, E)$ the chromatic number of G denoted $\chi(G)$ is the smallest k for which G has a k colouring.

Proposition 5.41. $\chi(K_n) = n$.

Proof. It follows that we can always find an n colouring for any graph with n vertices so it suffices to show that $\chi(K_n) > n - 1$. Suppose not and $\chi(K_n) \leq n-1$, then by the pigeon hole principle there are at least 2 vertices x and y of the same colour. However $xy \in E(K_n)$ and so x and y cannot have the same colour contradicting that $\chi(K_n) \leq n-1$. Thus it follows that $\chi(K_n) = n$. □

Proposition 5.42. *Let $G(V, E)$ be a graph, then*

$$\chi(G) \geq \omega(G).$$

Proof. Suppose that $\omega(G) = r$ then it follows that $K_r \subseteq G$. Thus

$$\begin{aligned} \chi(G) &\geq \chi(K_r) \\ &= r \end{aligned}$$

and the result follows. □

Proposition 5.43. *Let $G = (V, E)$ be a graph, then*

$$\chi(G) \geq \frac{|V(G)|}{\alpha(G)}.$$

Proof. Let $\chi(G) = k$ and consider an arbitrary k colouring of G. Since each of the k colour classes forms an independent set of vertices, let $V(i)$ denote the set of vertices in colour class $1 \leq i \leq k$. It follows that $|V(i)| \leq \alpha(G)$ and furthermore since every vertex is assigned a colour, the set of colour classes partitions the vertex set. Hence $V(G) = \bigcup_{i=1}^{k} V(i)$, and it follows that

$$\begin{aligned} |V(G)| &= \sum_{i=1}^{k} |V(i)| \\ &\leq \sum_{i=1}^{k} \alpha(G) \\ &= k\alpha(G) \\ &= \chi(G)\alpha(G) \end{aligned}$$

thus the result follows. □

Proposition 5.44. *For every graph $G = (V, E)$ with $|E| = m$ it follows that*

$$\chi(G) \leq 1/2 + \sqrt{(2m + 1/4)}.$$

Proof. Let $\chi(G) = k$ and consider an arbitrary k-colouring of G. For every pair of colour classes we must have at least one edge between them, otherwise we could colour both of the classes with the same colour using one less colour to colour the graph. Hence

$$\begin{aligned} m &\geq \binom{k}{2} \\ &= \frac{k(k-1)}{2}. \end{aligned}$$

Solving for k we obtain $k \leq 1/2 + \sqrt{(2m + 1/4)}$. □

Proposition 5.45. *Given a graph* $G = (V, E)$

$$\chi(G) \leq \Delta(G) + 1.$$

Proof. We use a greedy colouring algorithm to colour the vertices of G with colours from the set $\{1, 2, \ldots, \Delta(G) + 1\}$. Arbitrarily enumerate of the vertices of G as $v_1, \ldots, v_n$ and colour them in the following way. Colour v_1 with colour 1 and for each vertex v_i $(i > 1)$ colour it using the least colour available from the set $\{1, 2, \ldots, \Delta(G) + 1\}$ not already used to colour one if its neighbours in $\{v_1, \ldots, v_{i-1}\}$. In this way we never use more than $\Delta(G) + 1$ colours as each vertex has at most $\Delta(G)$ vertices preceding it in the enumeration. □

Definition 5.46. Define the colouring number denoted $\text{col}(G)$ of a graph $G = (V, E)$ to be least number k such that there exists an enumeration of the vertices in which each vertex is preceded by fewer than k of it's neighbours.

Proposition 5.47. $\text{col}(G) = \max\{\delta(H)|H \subseteq G\} + 1$.

Proof. We first show that $\text{col}(G) \leq \max\{\delta(H)|H \subseteq G\} + 1$, by showing that we can construct an enumeration$v_1, v_2, \ldots, v_n$ of the vertices such that each v_i for $(1 \leq i \leq n)$ has less than $\max\{\delta(H)|H \subseteq G\} + 1$ neighbours preceding it in the enumeration. Start by selecting the vertex v_n such that $d(v_n) = \delta(G)$. Next select the vertex v_{n-1} to be the vertex of minimum degree in $G_{n-1} = G[V \setminus \{v_n\}]$ and so on, thus v_i is the vertex of minimum degree in the induced subgraph $G_i = G[V \setminus \{v_{i+1}, v_{i+2}, \ldots, v_n\}]$. It follows that v_i has at most $\max\{\delta(H)|H \subseteq G\}$ neighbours preceding it in the enumeration. Hence $\text{col}(G) \leq \max\{\delta(H)|H \subseteq G\} + 1$.
Furthermore for each $H \subset G$ clearly $\text{col}(G) \geq \text{col}(H)$ and furthermore $\text{col}(H) \geq \delta(H) + 1$ since the final vertex in any enumeration of the vertices of H will have all of it's neighbours preceding it. Hence $\text{col}(G) = \max\{\delta(H)|H \subseteq G\} + 1$.

□

Proposition 5.48. *For* $G = (V, E)$ *it follows that*

$$\chi(G) \leq col(G).$$

Proof. Use the enumeration $v_1, v_2, \ldots, v_n$ of the vertices obtained from Proposition 5.47. Colour the vertices greedily from the set $\{1, 2, \ldots, col(G)\}$ as follows. Colour v_1 with colour 1, and for each vertex v_i where $i \geq 1$, select the first available colour from the set $\{1, 2, \ldots, col(G)\}$ not already used to colour it's neighbours preeceding it in the enumeration. Since each v_i has at most $col(G) - 1$ neighbours preceeding it the enumeration, we require at most $col(G)$ colours to colour the vertices and the result follows. □

5.2.1 Triangle free graphs with arbitrarily large Chromatic number

Definition 5.49. From a simple graph G, Mycielski's construction produces a simple graph G' containing G. We obtain G' as follows. Beginning with G having a vertex set $\{v_1, v_2, \ldots, v_n\}$, add vertices $U = \{u_1, u_2, \ldots, u_n\}$ and one more vertex w. Add edges such that vertex u_i is adjacent to all vertices $N_G(v_i)$, and finally let $N(w) = U$.

Theorem 5.50 (**Mycielski [15]**). *From a k-chromatic triangle-free graph G, Mycielski's construction produces a $k+1$-chromatic triangle free graph G'.*

Proof. Let the vertex set of the graph G be $V(G) = \{v_1, v_2, \ldots, v_n\}$ and let G' be the graph produced by applying Mycielski's construction to it. Let the set $U = \{u_1, u_2, \ldots, u_n\}$ and vertex w be defined as in Definition 5.49 above. By construction U is an independent set in G'. Hence the other two vertices of any triangle containing u_i would belong to the set $V(G)$. Suppose there exists such a triangle in G' with the other two vertices being v_j, v_k. It follows from construction that there also exists a triangle with vertices v_i, v_j, v_k contradicting that G is triangle free. We conclude that G' is triangle free.
A k-colouring f of G can extend to a $k+1$ colouring of G' as follows. Let $f(u_i) = f(v_i)$ and set $f(w) = k+1$. We can easily verify this is a $k+1$ colouring of G' colouring, since if $v_j u_i$ is an edge of G' then $f(v_j) \neq f(v_i) = f(u_i)$. Thus $\chi(G') \leq \chi(G)+1$ and it remains to verify that $\chi(G) < \chi(G')$. To prove this we show that if G' could be coloured with k colours then we can obtain a colouring of G with fewer than k colours.

Let g be a k colouring of G' and wlog assume that $g(w) = k$. This then restricts g to $\{1, \ldots, k-1\}$ on U and all k colours are used on $V(G)$. Let A be the set of vertices in $V(G)$ which are coloured with colour k. We show we can change the colours in A to obtain a $k-1$ colouring of G. For each vertex $v_i \in A$ change the colour of v_i to the colour $g(u_i)$. Since A is an independent set of vertices to verify that this procedure gives a $k-1$ colouring of G, it remains to check that if $v_i v_j$ is an edge in G where $v_i \in A$ and $v_j \in V(G) \setminus A$ that v_i and v_j receive different colours. This holds since the new colour of v_i is $g(u_i) \neq g(v_j)$. Thus our modified colouring gives a $k-1$ colouring of G. □

5.3 Ramsey Theory

In this section we investigate what kind of substructures are necessarily present in every large enough graph. For example how large will our graph have to be to necessarily contain a copy of K_r or a set of r independent vertices?

Example 5.51. Given a room with at least 6 people there are at least 3 who all know each other, or 3 of which no two know each other.

Proof. Assume that our room contains exactly 6 people (otherwise choose a group of 6 people) and denote them by $x_1, x_2, \ldots, x_6$. Person x_1 either knows at least 3 of the other 5 people or does not know at least three of the other 5 people. Wlog assume that person x_1 knows at least 3 of the other 5 people and assume three of these people are x_2, x_3 and x_4. Either no two of these people know each other or there is a pair which do. In either case we have a group of three people no two of which know each other or a group of three people who know each other. □

The same cannot be said about a group of 5 people as we can construct a scenario where for every set of three people there is exactly one pair which know each other or do not know each other. Construct a graph with vertices $x_1, x_2, \ldots, x_5$ representing our 5 people. Join an edge between two vertices x_i and x_j $(1 \leq i \neq j \leq 5)$ if person i and person j know each other. In the scenario where our graph is the cycle $C = x_1 x_2 \ldots x_5 x_1$, we have a construction where among any 3 people at least one pair will no each other but all

three do not know each other.

We now look at a graph theoretic question analogous to the example shown above. What is the minimum number of vertices required in a graph to ensure it either contains a triangle or an independent set of size 3? The answer is 6, and the proof is essentially given above. An equivalent formulation of the problem is as follows, if we were to colour each edge of K_n red or blue what is the smallest n such that no matter how we colour the edges we will have either a red triangle or blue triangle?

Definition 5.52. Given positive integers s and t the Ramsey number $R(s,t)$ is the order of the smallest complete graph, which when it's edges are coloured with two colours (red and blue), it must contain a red K_s or a blue K_t.

Claim 5.53. *The following simple results hold for ramsey numbers.*

1. $R(s,t) = R(t,s)$.
2. $R(s,1) = 1$.
3. $R(s,2) = s$.

We have shown that $R(3,3) = 6$ so let us know look at the Ramsey number $R(4,3)$.

Proposition 5.54. $R(4,3) = 9$.

Proof. Let us start by showing that $R(4,3) > 8$. Colour the edges of the labeled graph K_8 red and blue as follows. Let the set of blue edges be the set $E_B = \{12, 23, 34, 45, 56, 67, 78, 81, 15, 26, 37, 48\}$ and colour all other edges of K_8 red. It's not hard to see this graph does not contain a red K_4 or a blue K_3.
Next we show that $R(4,3) \leq 9$. Label the vertices of K_9, 1-9. There must be some vertex $x \in \{1, 2, \ldots, 9\}$ incident to at least 6 red edges or at least 4 blue edges. Otherwise each vertex would be incident to exactly 5 red edges and 3 blue edges and we would obtain a red subgraph in which all degrees are odd. However Proposition 5.14 tells us that the sum of the degrees of the vertices in any graph must be even.
Suppose x is incident to 6 red edges, then since $R(3,3) = 6$ the graph on these 6 vertices must contain a blue K_3 or a red K_3. If it contains a blue K_3 our graph contains a blue K_3. On the other hand if it contains a red K_3

together with the edges from x our graph contains a red K_4.
Suppose instead that x is incident to 4 blue edges. If for any pair of these 4 vertices the edge connecting them is blue together with edges from x we have a blue K_3. If not, all edges between these 4 vertices must be coloured red and hence our graph contains a red K_4. Thus in either case the graph contains a red K_4 or a blue K_3. □

Theorem 5.55 (**Ramsey's Theorem [18]**). *For any two natural numbers, s and t, there exists a natural number $R(s,t) = n$ such whenever the edges of K_n are coloured red or blue it must contain a red K_s or blue K_t.*

We prove that $R(s,t)$ exists for every $s,t \in \mathbb{N}$. We proceed by induction first assuming that $R(s,t-1)$ and $R(s-1,t)$ exist.

Claim 5.56. $R(s,t) \leq R(s-1,t) + R(s,t-1)$.

Proof. We have already seen from claim 5.53 that $R(s,1) = R(1,s) = 1$ and $R(s,2) = R(2,s) = s$. For our inductive hypothesis suppose that $R(s-1,t)$ and $R(s,t-1)$ both exist. Let $n = R(s-1,t) + R(s,t-1)$ and consider any red-blue colouring of K_n.Pick an arbitrary vertex x and partition the remaining $n-1$ vertices into two disjoint sets R_x and B_x, with R_x being the set of vertices adjacent to x with red edges and B_x the set of vertices adjacent to x with blue edges.

Since $n = R(s-1,t) + R(s,t-1)$ either $|R_x| \geq R(s-1,t)$ or $|B_x| \geq R(s,t-1)$ since otherwise $|R_x| + |B_x| \leq n-2$ contradicting the fact that $|R_x| + |B_x| = n-1$.
If $|B_x| \geq R(s,t-1)$ and B_x induces a red K_s then we are done. Otherwise it induces a blue K_{t-1}, then K_n must contain a blue K_t as $B_x \cup \{x\}$ will induce a blue K_t (as each edge xy where $y \in B_x$ is blue). The case where $|R_x| \geq R(s-1,t)$ is symmetric and so we have shown by induction that any red-blue colouring of K_n must contain a red K_s or blue K_t. □

Corollary 5.57. $R(s,t) \leq \binom{s+t-2}{t-1}$.

Proof. We use induction on $s+t$. The result holds for all Ramsey numbers $R(s,2)$ and $R(2,t)$ where $s,t \in \mathbb{N}$ since

$$R(s,2) = s = \binom{s}{1} \quad \text{and} \quad R(2,t) = t = \binom{t}{t-1}.$$

Suppose the result holds for $R(s-1,t)$ and $R(s,t-1)$. It follows from Claim 5.56 that

$$\begin{aligned} R(s,t) &\leq R(s-1,t)+R(s,t-1) \\ &\leq \binom{s+t-3}{t-2}+\binom{s+t-3}{t-1} \\ &= \binom{s+t-2}{t-1} \end{aligned}$$

and the result follows. □

Chapter 6

Random Graphs

We have introduced the fundamental concepts of Graph Theory and Probability Theory. We now aim to combine the two by exploring the notion of random graphs. A random graph on n vertices is obtained by selecting a subset of $E(K_n)$ at random; how we select this subset depends on the model we are using. There are two standard models which we will look at in this chapter. The first of these models is the uniform model $G(n, m)$ proposed by Erdős and Réyni [10] and the second is the Binomial model $G(n, p)$ proposed by Gilbert [12]. In the uniform model we select $0 \leq m \leq \binom{n}{2}$ edges at random, and in this model each labeled graph with m edges has an equal chance of being selected. In the binomial model each edge of $E(K_n)$ is selected independently with probability $0 \leq p \leq 1$. We start by looking at the uniform model first and then the binomial model second. We will compare the two models once we have introduced them both.

6.1 The Uniform random graph model

The first model we study was proposed by Erdős and Réyni in 1959 in a seminal paper. As mentioned above, in the random graph model $G(n, m)$ a labeled graph on n vertices and m edges is chosen at random with each graph having equal probability of being selected. Thus $G(n, m)$ is a uniform probability space. We start by counting the number of graphs for fixed n and m in the uniform model. There are precisely $N = \binom{n}{2}$ edges of K_n. Each graph in $G(n, m)$ corresponds to a selection of m of these edges. It follows that there are $\binom{N}{m}$ graphs on n vertices with m edges.

Definition 6.1. Let G_n^m denote the set of all labeled graphs on n vertices with m edges. Then $G(n,m)$ is the probability space $G(n,m) = (G_n^m, \mathcal{P}(G_n^m), P)$ with $P(G(n,m) = G) = \frac{1}{\binom{N}{m}}$ (where $N = \binom{n}{2}$) for all $G \in G_n^m$.

Example 6.2. Consider $G(4,3)$ and let $G \in G(4,3)$ be the graph with edge set $E(G) = \{12, 24, 34\}$. Then

$$\begin{aligned} P(G(n,m) = G) &= \frac{1}{\binom{\binom{4}{2}}{3}} \\ &= \frac{1}{\binom{6}{3}} \\ &= \frac{1}{20}. \end{aligned}$$

Suppose we want to know something more interesting about the structure of $G(4,3)$, perhaps the probability that $G(4,3)$ is connected. Any graph on 4 vertices with three edges is connected if and only if it does not consist of a 3-cycle and an isolated vertex. Thus there are exactly 4 disconnected graphs on 4 vertices with 3 edges and so there are 16 connected graphs. Hence

$$P(G(4,3) \text{ is connected}) = \frac{16}{20}\ .$$

We now try and generalize the result in Example 6.2 by analyzing what happens to the probability of a graph being connected if it contains exactly half of all it's potential edges as $n \to \infty$.

Proposition 6.3. *$G(n,m)$ is connected a.a.s when $m = N/2$ where $N = \binom{n}{2}$.*

Proof. We wish to count the number of connected graphs on n labeled vertices with exactly m edges. Let C_n^m denote the set of all such graphs. Then $P(G(n,m) \in C_n^m) = \frac{|C_n^m|}{\binom{N}{m}}$. Similar to Example 6.2 it is easier for us count the number of disconnected graphs in order for us to find the number of connected ones.

To count the number of disconnected graphs on n vertices and m edges, we note that the vertices of such a graph can be partitioned into two disjoint vertex classes with no edges between them. Call a partitioning of this form a graph partition. It is important to note that it is possible for a graph to have more than one graph partition. For example suppose $G = (V, E)$ has a graph partition into vertex classes A and B with $|A| \leq |B|$, and in addition suppose A contains an isolated vertex x. Then $G = (V, E)$ has at least one other graph partition with classes $A' = A \setminus \{x\}$ and $B' = B \cup \{x\}$. Furthermore fixing the class A of a partition also fixes the class B as each vertex not assigned to A is assigned to B.

Suppose we consider all disconnected graphs which have a graph partition into fixed vertex classes A and B, of sizes a and b respectively where $a \leq b$. There are exactly $\binom{N-ab}{m}$ such graphs, each containing no edges between the vertex classes A and B. Furthermore there are $\binom{n}{a}$ ways to choose the labels of the vertices in the class A. Hence so there are exactly $\binom{n}{a}\binom{N-ab}{m}$ graph partitions. If we now sum $\binom{n}{a}\binom{N-ab}{m}$ over all possible sizes of the set A we obtain an upper bound on the number of disjoint graphs on n vertices with m edges.

Let D_n^m denote the set of all disconnected graphs on n vertices with m edges. Then it follows that

$$\begin{aligned} |D_n^m| &\leq \sum_{a=1}^{n/2} \binom{n}{a}\binom{N-ab}{m} \\ &= \sum_{a=1}^{n/2} \binom{n}{a}\binom{N-a(n-a)}{m} \end{aligned}$$

with the equality following from the fact that $a + b = n$.
We will use the following results to analyze $P(G(n, m) \in D_n^m)$ as $n \to \infty$.

$$\frac{\binom{N-a(n-a)}{m}}{\binom{N}{m}} \leq e^{-\frac{N}{m}a(n-a)} = e^{-2a(n-a)} \text{ as } m = N/2. \tag{6.1}$$

$$\binom{n}{a} \leq \binom{n}{n/2} \leq \left(\frac{ne}{n/2}\right)^{n/2} = (2e)^{n/2}\,. \tag{6.2}$$

$$e^{-2a(n-a)} \leq e^{-2n} \text{ for } a > 1 \text{ and large } n. \tag{6.3}$$

We can now bound $P(G(n,m) \in D_n^m)$ as follows,

$$\begin{aligned}
P(G(n,m) \in D_n^m) &\leq \frac{\sum_{a=1}^{n/2} \binom{n}{a}\binom{N-a(n-a)}{m}}{\binom{N}{m}} \\
&= \frac{\binom{n}{1}\binom{N-(n-1)}{m}}{\binom{N}{m}} + \frac{\sum_{a=2}^{n/2} \binom{n}{a}\binom{N-a(n-a)}{m}}{\binom{N}{m}} \\
&\leq \frac{\binom{n}{1}\binom{N-(n-1)}{m}}{\binom{N}{m}} + \frac{\sum_{a=2}^{n/2} \binom{n}{n/2}\binom{N-a(n-a)}{m}}{\binom{N}{m}} \\
&\leq ne^{-2(n-1)} + \sum_{a=2}^{n/2} (2e)^{n/2} e^{-2a(n-a)} \\
&\leq ne^{-2(n-1)} + \sum_{a=2}^{n/2} (2e)^{n/2} e^{-2n} \\
&\leq ne^{-2(n-1)} + n/2 \cdot (2e)^{n/2} e^{-2n} \\
&= ne^{-2(n-1)} + ne^{-3n/2} 2^{n/2-1} \\
&= n(e^{-2(n-1)} + e^{-3n/2} 2^{n/2-1}) \\
&= o(1).
\end{aligned}$$

The second inequality follows from (4.2) the third inequality follows from (4.1) and (4.2), finally the fourth inequality follows from (4.3).
Since $P(G(n,m) \in D_n^m) = o(1)$ we conclude that $G(n,m)$ is connected a.a.s when $m = N/2$.

6.2 Binomial Random graph model

One of the issues that occur in the $G(n,m)$ random graph model is that edges do not occur independently of one another. This often makes calculations rather difficult as seen in the proof of Proposition 6.3. We now look at the Binomial random graph model $G(n,p)$, which is a lot easier to work with. In this model all edges are selected independently of one another.

Definition 6.4. Let G^n be the set of all labeled graphs on n vertices and let $0 \leq p = p(n) \leq 1$. Then the binomial random graph model $G(n,p)$ is the

probability space $(G^n, \mathcal{P}(G^n), P)$ with

$$P(G(n,p) = G) = p^{|E(G)|}(1-p)^{\binom{n}{2}-|E(G)|}$$

for all $G \in G^n$.

Proposition 6.5. *The expected number of edges in $G(n, 1/2)$ is $N/2$ where $N = \binom{n}{2}$.*

Proof. Enumerate the edges of K_n from 1 to N. For edge $1 \leq i \leq N$ let X_i be the indicator random variable such that

$$X_i = \begin{cases} 1, \text{ if edge } i \text{ is an edge in } G(n,p) \\ 0, \text{ otherwise .} \end{cases}$$

The random variable $X = \sum_{i=1}^{N} X_i$ counts the total number of edges in our random graph. Furthermore

$$\begin{aligned} \mathbb{E}(X) &= \sum_{i=1}^{N} \mathbb{E}(X_i) \\ &= \sum_{i=1}^{N} P(X_i = 1) \\ &= \sum_{i=1}^{N} 1/2 \\ &= N/2, \end{aligned}$$

the first equality follows from the Linearity of Expectation and the second equality follows from Theorem 4.3. □

Example 6.6. Consider an analogous case to Example 6.2 for $G(n,p)$ by considering the probability that $G(4, 1/2)$ is connected. Any graph on 4 vertices with less than 3 edges is disconnected. As mentioned in Example 6.2 a graph on 4 with 3 edges is disconnected if an only if it contains a 3-cycle

and an isolated vertex. Finally any graph on 4 vertices with more than 3 edges is connected. Hence it follows that

$$\begin{aligned} P(G(4,1/2) \text{ is disconnected}) &= \binom{6}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^6 + \binom{6}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^5 \\ &+ \binom{6}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^4 + 4\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^3 \\ &= \left(\sum_{i=0}^{2}\binom{6}{i} + 4\right)\left(\frac{1}{2}\right)^6 \\ &= \frac{26}{64}. \end{aligned}$$

Finally

$$\begin{aligned} P(G(4,1/2) \text{ is connected}) &= 1 - \frac{25}{64} \\ &= \frac{38}{64}. \end{aligned}$$

We now prove an analogous result to Proposition 6.3 for $G(n,p)$.

Theorem 6.7. *$G(n,1/2)$ is connected a.a.s*

Proof. We prove a stronger claim, namely that between any two vertices $i,j \in V(G)$ there is a path of length at most 2 a.a.s in $G(n,1/2)$. Start by enumerating the vertices of $G(n,1/2)$ from 1 to n. Define X_{ij} to be the indicator random variable for vertices $1 \leq i < j \leq n$ such that

$$X_{ij} = \begin{cases} 1, \text{ if there is no path of length atmost 2 between } i \text{ and } j \\ 0, \text{ otherwise .} \end{cases}$$

Let X be the random variable that counts the number of pairs of vertices which do not have a path of length at most 2 between them. Then

$$X = \sum_{1 \leq i < j \leq n} X_{ij}.$$

Observe that for any vertices i,j if there is a path of length atmost 2 between i and j then either there exists an edge between i and j or they must have a

common neighbour. Thus $P(X_{ij} = 1) = \frac{1}{2}(\frac{3}{4})^{n-2}$. This is because the probability that there does not exist an edge between i and j is $\frac{1}{2}$ and for each of the $n-2$ vertices $k \in V(G) \setminus \{i, j\}$ the probability that $k \notin N(i) \cap N(j)$ is $\frac{3}{4}$ and the appearance of edges are independent of one another.

The expectation of X is given by

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{1 \leq i < j \leq n} \mathbb{E}(X) \\
&= \sum_{1 \leq i < j \leq n} \frac{1}{2} \left(\frac{3}{4}\right)^{n-2} \\
&= \binom{n}{2} \frac{1}{2} \left(\frac{3}{4}\right)^{n-2} \\
&= o(1).
\end{aligned}$$

It follows that $P(X > 0) \leq \mathbb{E}(X)$ from Markovs inequality and thus $X = 0$ a.a.s. Hence a.a.s for every pair $i, j \in V(G)$ there is a path of length at most two between i and j and so it follows that $G(n, \frac{1}{2})$ is connected a.a.s. □

It is evident that this proof is much easier than the one used to prove its analogous counterpart for $G(n, N/2)$. We tend to therefore work with $G(n, p)$ rather than $G(n, M)$, as calculations become a lot simpler. Later on we will see how the two models appear to be very similar when $p = \frac{M}{N}$. In this case both models have the same expected number of edges.

6.3 Random Graph Process

Let $\sigma : E(K_n) \to [N]$ where $N = \binom{n}{2}$ be a permutation of the set of edges of K_n; we denote $\sigma = (e_1, e_2, \ldots, e_N)$. A graph process $G(\sigma)$ is a sequence of labeled graphs $(G_i)_{i=0}^{N}$ on n vertices where $G_0 = ([n], \emptyset)$ and the edge set of G_i is $\{e_1, e_2, \ldots, e_i\}$ (note the vertex set of each G_i is $[n]$). Hence the sequence starts with the empty graph on n vertices and ends with the complete graph on n vertices. In each subsequent step in moving from i to $i+1$, we add a single edge namely e_{i+1} to the graph G_i to obtain the graph G_{i+1}. Thus the graph G_i is contained within the graph G_{i+1} and so $G(\sigma)$ is a nested sequence of graphs. Moreover we can think of a graph process as an evolution of the

empty graph to the complete one guided by σ. A random graph process takes an empty graph and slowly evolves it into a complete graph by choosing σ uniformly at random from the set of all permutations of $[N]$. This process is very useful of this as it allows us to try and determine the point in the evolution that the graph first satisfies certain desired properties.

Definition 6.8. Let $\sigma = (e_1, e_2, \ldots, e_N)$ be a permutation of the edges of K_n, $G_0 = ([n], \emptyset)$ and $G_i = ([n], (e_1, e_2, \ldots, e_i)$ for all $1 \leq i \leq N$. When σ is chosen uniformly at random we call $G(\sigma)$ a random graph process.

Proposition 6.9. *A random graph process $G(\sigma)$ is equivalent to starting with an empty graph and at each stage selecting a edge uniformly at random from all possible edges not already selected.*

Proof. Given a permutation of the edges of K_n $\sigma = (e_1, e_2, \ldots, e_N)$, the probability that we select σ by selecting one edge at a time until all edges have been selected is

$$\frac{1}{N} \cdot \frac{1}{N-1} \cdot \ldots \cdot \frac{1}{2} \cdot \frac{1}{1} = \frac{1}{N!}.$$

The probability that we select σ by selecting one permutation of the edges of K_n uniformly at random is also $\frac{1}{N!}$ as there are a total of $N!$ permutations. Hence the two methods of selection are equivalent. □

One of the reasons that a random graph process is very important is because if we stop the process or take a snapshot at time j we produce the uniform probability distribution $G(n, j)$. Therefore understanding a random graph process leads to immediate consequences for $G(n, m)$.

Proposition 6.10. *If $G = G(\sigma)$ is a random graph process. Then $G_i \sim G(n, i)$*

Proof. Consider an arbitrary graph $G = ([n], E)$ with $|E| = i$. We wish to compute the probability that at stage i in our random graph process the graph we obtain is isomorphic to G. As long as the first i edges that appear are the edges in G then we have $G_i = G$. There are $i!$ permutations of these i edges and we do not care about the order in which the remaining edges appear. Hence there are exactly $i!(N-i)!$ permutations for which $G_i = G$.

Thus $P(G_i = G) = \frac{i!(N-i)!}{N!} = \frac{1}{\binom{N}{i}}$ and so it follows from Definition 6.1 that $G_i \sim G(n, i)$. □

6.4 Staged exposure

It is often useful to combine two or more disjoint random graphs to obtain another random graph. The idea of staged exposure shows us how this is possible in the context of the binomial random graph.

Proposition 6.11. *Suppose* $0 \leq p_1 \leq p_2 \cdots \leq p_k \leq 1$ *and further that* $1 - p = (1 - p_1)(1 - p_2) \cdots (1 - p_k)$ *then the probability distributions* $G(n, p)$ *and* $G' = \bigcup_{i=1}^{k} G(n, p_i)$ *are identical.*

Proof. Arbitrarily enumerate the edges of K_n as e_1 to e_N. For edge e_i where $1 \leq i \leq N$ it follows that $P(e_i \in G(n, p)) = p$ by Definition 6.4. Furthermore e_i is an edge of G' if it appears in at least one of the random graphs $G(n, p_i)$ for $1 \leq i \leq k$. Hence

$$\begin{aligned} P(e_i \in G') &= 1 - P(e_i \notin G') \\ &= 1 - (1 - p_1)(1 - p_2) \cdots (1 - p_k) \\ &= 1 - (1 - p) \\ &= p \, . \end{aligned}$$

In both distributions the probability that each edge appears independently of others is equal to p and so the two distributions are equal. □

6.4.1 Monotonicity

Proposition 6.12. *Let* $\mathcal{P}$ *be a monotone increasing graph property and* $0 \leq m_1 \leq m_2 \leq N$. *Then*

$$P(G(n, m_1) \in \mathcal{P}) \leq P(G(n, m_2) \in \mathcal{P}).$$

Proof. Let $(\sigma_S, \mathcal{P}(\sigma_S), P)$ be a probability space with σ_S the set of all $N!$ permutations $\sigma : E(K_n) \to [N]$ where $N = \binom{n}{2}$ and $P(\omega) = \frac{1}{N!}$ for all $\omega \in \sigma_S$. Using Proposition 6.10 it follows that for ever $0 \leq i \leq N$ the snapshot at time i models $G(n, i)$.

Given an increasing monotone graph property $\mathcal{P}$, let A be the event $G_{m_1} \in \mathcal{P}$ and B the event $G_{m_2} \in \mathcal{P}$ where $G_{m_1} \sim G(n, m_1)$ and $G_{m_2} \sim G(n, m_2)$. If a permutation σ satisfies the event A then because $\mathcal{P}$ is a monotone increasing graph property; it follows that σ also satisfies the event B. Hence $A \subseteq B$ and it follows that $P(A) \leq P(B)$. □

Proposition 6.13. *Let P be a monotone graph property and $0 \leq p_1 \leq p_2 \leq 1$. Then*

$$P(G(n, p_1) \in \mathcal{P}) \leq P(G(n, p_2) \in \mathcal{P}).$$

Proof. Let $G_1 \sim G(n, p_0)$, $G_2 \sim G(n, p_2)$ and choose $G_0 \sim G(n, p_0)$ such that $(1-p_0)(1-p_1) = (1-p_2)$. Since $0 \leq \frac{1-p_2}{1-p_1} \leq 1$ it follows that $0 \leq 1-\frac{1-p_2}{1-p_1} \leq 1$ and thus p_0 exists. By Proposition 6.11 we can represent G_2 by $G_0 \cup G_1$. Consider the event A that $G(n, p_1) \in \mathcal{P}$ and the event B that $G(n, p_2) \in \mathcal{P}$. It follows that if $G_1 \in \mathcal{P}$ then $G_2 \in \mathcal{P}$ as $G_2 = G_1 \cup G_0$ and so $A \subseteq B$. Finally we conclude that $P(A) \leq P(B)$. □

6.5 Two important proofs which use Random Graphs

We now look at two classic examples which use random graphs to prove theorems in combinatorics.

Theorem 6.14 (**Erdős [11]**). *$R(k, k) > 2^{k/2}$ for k large enough.*

Proof. It suffices to show that there exists a graph $G = ([n], E)$ with $n = \lfloor 2^{k/2} \rfloor$ such that $\omega(G) < k$ and $\alpha(G) < k$. Consider $G(n, 1/2)$ where $n = \lfloor 2^{k/2} \rfloor$. Arbitrarily enumerate all sets of k vertices from 1 to $\binom{n}{k}$ and let X_i be the indicator random variable for the set $1 \leq i \leq \binom{n}{k}$ such that

$$X_i = \begin{cases} 1, \text{ if } i \text{ is a } k-\text{clique} \\ 0, \text{ otherwise .} \end{cases}$$

Let X be the random variable which represents the total number of k cliques in G. Thus $X = \sum_{i=1}^{\binom{n}{k}} X_i$.

Given the ith set of k vertices, we find $P(X_i = 1) = (\frac{1}{2})^{\binom{k}{2}}$ and so it follows that

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{i=1}^{\binom{n}{k}} \mathbb{E}(X_i) \\
&= \sum_{i=1}^{\binom{n}{k}} \left(\frac{1}{2}\right)^{\binom{k}{2}} \\
&= \binom{n}{k}\left(\frac{1}{2}\right)^{\binom{k}{2}} \\
&\leq \left(\frac{en}{k}\right)^k 2^{-k(k-1)/2} \\
&= \left(\frac{en}{k}2^{-\frac{k-1}{2}}\right)^k \\
&\leq \left(\frac{e2^{k/2}}{k}2^{-k/2+1/2}\right)^k \\
&= \left(\frac{e2^{1/2}}{k}\right)^k \\
&< 1/2 \ ,
\end{aligned}$$

with the first inequality following from the linearity of expectation and the second inequality holds because $n = \lfloor 2^{k/2} \rfloor$, (note the final inequality holds for large k). Using Markov's inequality we obtain $P(X > 0) \leq \frac{\mathbb{E}(X)}{1} < 1/2$.

Let Y be the number of independent sets of size k in G. Then by the same computation as above

$$\mathbb{E}(Y) = \binom{n}{k}\left(\frac{k}{2}\right)^{\binom{k}{2}} < 1/2$$

and it follows from Markov's inequality that $P(Y > 0) \leq \mathbb{E}(Y) < 1/2$.

Finally $P(X = 0 \text{ and } Y = 0) = 1 - P(X \geq 1 \text{ or } Y \geq 1)$. Using the union bound we obtain $P(X \geq 1 \text{ or } Y \geq 1) \leq P(X \geq 1) + P(Y \geq 1) < 1$. Hence $P(X = 0 \cap Y = 0)$ has a positive probability and it follows that there exists a graph G with on n vertices with $\omega(G) < k$ and $\alpha(G) < k$. □

We now look at an interesting result which combines the girth and chromatic number of a graph. It is intuitive to assume that the larger the girth of a graph the less edges it has. Using the bound from Proposition 5.44 we would then expect the chromatic number to be bounded above. Contrary to this Erdős showed that it is possible using the theory of random graphs to find graph with high girth and high chromatic number using random graph.

Theorem 6.15 (**Erdős [9]**). *For all integers $k, l \geq 3$ there is a graph G with $g(G) > l$ and $\chi(G) > k$.*

Proof. Start by fixing a constant θ with $0 < \theta < 1/l$ and consider the random graph $G(n, p)$ with $p = n^{-1+\theta}$. Let X be the random variable representing the number of cycles of length at most l in G. We are interested in calculating the expected number of i-cycles for all $i \leq l$ in G. In order to do this we start by calculating exactly how many i-cycles G contains for each $i \leq l$.

Consider all ordered i-tuples of vertices for $i \leq l$. Let the i-tuple

$$(v_1, v_2, \cdots, v_i)$$

correspond to the cycle $C = v_1 v_2 \cdots v_i v_1$. Then the i tuples

$$(v_1, v_2, \cdots, v_i), (v_2, v_3, \ldots, v_1), (v_3, v_4, \ldots, v_2), \ldots, (v_i, v_1, \ldots, v_{i-1}),$$

including the i-tuples obtained by reversing the order of the ones presented all correspond to the same cycle C. Thus there are precisely $2i$ ordered i-tuples which all correspond to the same cycle. Let $c(i)$ denote the number of i-cycles for $i \leq l$, it follows that

$$c(i) = \frac{n(n-1)\ldots(n-(i-1))}{2i}.$$

For each ordered i-tuple of vertices the probability that G contains the specified cycle on those i vertices is p^i.

The expected number of i cycles for $i \leq l$ is then

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{i=3}^{l} c(i)p^i \\
&= \sum_{i=3}^{l} \frac{n(n-1)\cdots(n-(i-1))}{2i} p^i \\
&\leq \sum_{i=3}^{l} n^i p^i \\
&= \sum_{i=3}^{l} n^{\theta i}
\end{aligned}$$

the last equality follows because $p = n^{-1+\theta}$. Since

$$\mathbb{E}(X) \leq \sum_{i=3}^{l} n^{\theta i} = O(n^{\theta l}) = o(n)$$

it follows from Markov's inequality that

$$P(X \geq n/2) \leq \frac{\mathbb{E}(X)}{n/2} = o(1)$$

and hence

$$P(X \geq n/2) = o(1).$$

Next we bound the clique number $\alpha(G)$. Start by enumerating all sets of vertices of size t from 1 to $\binom{n}{t}$. For set $1 \leq i \leq \binom{n}{t}$ let Y_i be the indicator random variable such that

$$Y_i = \begin{cases} 1, \text{ i is an independent set} \\ 0, \text{ otherwise .} \end{cases}$$

For set i because each of the $\binom{t}{2}$ potential edges have probability $1-p$ of not being selected it follows that

$$P(Y_i = 1) = (1-p)^{\binom{t}{2}}.$$

Let Y be the random variable which represents the total number of independent sets of size t, then

$$Y = \sum_{i=1}^{\binom{n}{t}} Y_i.$$

The expectation of Y is then given by

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{i=1}^{\binom{n}{t}} \mathbb{E}(Y_i) \\
&= \sum_{i=1}^{\binom{n}{t}} (1-p)^{\binom{t}{2}} \\
&= \binom{n}{t}(1-p)^{\binom{t}{2}}.
\end{aligned}$$

If we set $t = \lceil \frac{3\ln(n)}{p} \rceil$ it follows that

$$\begin{aligned}
\mathbb{E}(Y) &= \binom{n}{t}(1-p)^{\binom{t}{2}} \\
&= \left(\frac{en}{t}\right)^t e^{-1/2pt(t-1)} \\
&= \left(\frac{en}{t}e^{-1/2p(t-1)}\right)^t \\
&\leq \left(ene^{-1/2p(t-1)}\right)^t \\
&= \left(ene^{-1/2p(\lceil \frac{3\ln(n)}{p} \rceil - 1)}\right)^t \\
&\leq \left(ene^{-1/2p(2.8)\frac{\ln(n)}{p}}\right)^t \\
&= \left(ene^{-1.4\ln(n)}\right)^t \\
&= o(1).
\end{aligned}$$

Note that $\alpha(G) \geq t$ if and only if $Y > 0$ and hence

$$P(\alpha(G) \geq t) = o(1).$$

Combining the bounds obtained for random variables X and Y it follows that

$$\begin{aligned} P(Y > 0 \text{ or } X \geq n/2) &\leq P(Y > 0) + P(X \geq n/2) \\ &= o(1). \end{aligned}$$

Hence there exists a graph G on n vertices such that $\alpha(G) < t$ and $X(G) < \frac{n}{2}$. Next form an induced subgraph $G' \subseteq G$ as follows. For each cycle of length less than l in G delete an arbitrary vertex, thus $g(G') > l$. The number of vertices in G' satisfies $V(G') \geq \frac{n}{2}$ as there are less than $n/2$ cycles in G and so less than $n/2$ vertices were deleted in obtaining G'. Finally $\alpha(G') \leq \alpha(G) < t$ and so

$$\begin{aligned} \chi(G') &\geq \frac{|V(G')|}{\alpha(G')} \\ &\geq \frac{n/2}{\lceil \frac{3\ln(n)}{p} \rceil} \\ &\geq n^{\theta/2} \\ &\geq k. \end{aligned}$$

G' is then a graph with all the desired properties. □

6.6 Properties of almost all graphs

Given a graph property $\mathcal{P}$ we would often know how $P(G(n,p) \in \mathcal{P})$ behaves as $n \to \infty$, as this tells us roughly what proportion of graphs will contain the given property. If $P(G(n,p) \in \mathcal{P}) \to 1$ as $n \to \infty$ we say that almost all graphs contain property $\mathcal{P}$. We will illustrate some properties which are contained by almost all graphs.

Proposition 6.16. *For every constant $p \in (0,1)$ and every graph H, a.a.s $G \sim G(n,p)$ contains an induced copy of H.*

Proof. Let H be given, and suppose $k = |H|$. Let $n \geq k$ and suppose we select $U \subseteq \{1, \ldots, n\}$ of size k. Then $P(G[U])$ is isomorphic to H has probability $0 < r < 1$ with $r = r(p)$. Furthermore G contains $\lfloor \frac{n}{k} \rfloor$ disjoint sets of size k each with probability $(1-r)$ of not containing an induced subgraph

isomorphic to H independent of each other. Furthermore if a graph does not contain an induced copy of H then neither does any one if it's $\lfloor \frac{n}{k} \rfloor$ disjoint sets of size k. Hence

$$P(H \not\subseteq G \text{ induced}) \leq (1-r)^{\lfloor \frac{n}{k} \rfloor} \to 0$$

as $n \to \infty$. □

Given $i, j \in \mathbb{N}$, let $\mathcal{P}_{i,j}$ be the graph property that a given graph contains for any disjoint vertex sets U, W with $|U| \leq i$ and $|W| \leq j$, a vertex $v \notin U \cup W$ that is adjacent to all the vertices in U but to none in W.

Theorem 6.17. *For every constant $p \in (0,1)$ and $i, j \in \mathbb{N}$ almost every graph has the property $\mathcal{P}_{i,j}$. a.a.s*

Proof. For fixed disjoint sets U, W consider $v \in G \setminus (U \cup W)$, the probability that v is adjacent to all the vertices in U but to non in W is

$$p^{|U|}(1-p)^{|W|} \leq p^i(1-p)^j.$$

If we now consider all possible choices of the vertex v we find the probability that our graph does not contain the property $\mathcal{P}_{i,j}$ is

$$(1 - p^{|U|}(1-p)^{|W|}))^{n-|U|+|W|} \leq (1 - p^i(1-p)^j)^{n-i-j}$$

where $n \geq i + j$.
We obtain an upper bound on the number of possible selections for the sets $|U|$ and $|W|$. We have at most i vertices available for U and j vertices available for set W and hence $i + j$ positions in total. Consider a set of $i + j$ boxes numbered 1 to $i + j$. In each box place one of the n possible vertices allowing repetition. Let the set of vertices (not including repetition) of the first i boxes form the set U and the remaining j boxes form the set W. If do not worry about U and W being disjoint it follows that there are at most n^{i+j} possible selections for the sets $|U|$ and $|W|$. Thus the probability that there exists a pair of sets U, W such that no vertex v with the desired properties exists, is at most

$$n^{i+j}(1 - p^i(1-p)^j)^{n-i-j} \to 0$$

as $n \to \infty$. □

Corollary 6.18. *For every constant $p \in (0,1)$ and $k \in \mathbb{N}$, almost every graph in $G(n,p)$ is $k-$connected.*

Proof. Using Theorem 6.17 we know almost every graph has property $\mathcal{P}_{2,k-1}$. Let $G = ([n], E)$ be a graph which has the property $\mathcal{P}_{2,k-1}$ and suppose G is not k-connected. Then there exists a set W with $|W| \leq k-1$ such that the induced subgraph $G' = G[[n] \setminus W]$ is disconnected. Thus there must exist vertices x, y in G' with no path connecting them. To show this cannot be the case let $U = \{x, y\}$ and take W already given. Then since G has property $\mathcal{P}_{2,k-1}$ there exists a vertex v outside W such that v is a neighbor of both x and y. However xwy then forms a path in G', contradicting that G is not k-connected. □

We now present a lower bound on the chromatic number of almost every graph.

Proposition 6.19. *For every constant $p \in (0,1)$ and every $\epsilon > 0$, almost every graph $G \in G(n,p)$ has chromatic number*

$$\chi(G) > \frac{\log\left(\frac{1}{(1-p)}\right)}{2+\epsilon} \cdot \frac{n}{\log(n)} .$$

Proof. For any fixed $n \geq k \geq 2$,

$$\begin{aligned} P(\alpha \geq k) &\leq \binom{n}{k}(1-p)^{\binom{k}{2}} \\ &\leq n^k (1-p)^{\binom{k}{2}} \\ &= (1-p)^{k\frac{\log(n)}{\log(1-p)} + \frac{1}{2}k(k-1)} \\ &= (1-p)^{\left(\frac{k}{2}\left(-\frac{2\log(n)}{\log(1/(1-p))}\right)+k-1\right)}. \end{aligned}$$

If we let $k = (2+\epsilon)\frac{\log(n)}{\log(1/(1-p))}$ the exponent in this expression tends to infinity as $n \to \infty$ and so the entire expression tends to 0. Since almost every graph in $G \sim G(n,p)$ will have a maximum independent set of size less than k,

each colour class of our colouring will contain less than k vertices. Hence our colouring will use more than

$$\frac{n}{k} = \frac{\log(1/(1-p))}{2+\epsilon} \cdot \frac{n}{\log(n)}$$

colours. □

6.7 Asymptotic equivalence

When working with $G(n,p)$, if we set $p = m/\binom{n}{2}$ then the expected number of edges in our binomial random graph will be m. In this instance $G(n,p)$ and $G(n,m)$ the same number of expected edges. We will study the equivalence of the models when p is close to $m/\binom{n}{2}$. We will prove that for any given monotone graph property $\mathcal{P}$, the convergence of $P(G(n,m) \in \mathcal{P})$ implies the convergence of $P(G(n,p) \in \mathcal{P})$ to the same limit. We start by introducing some notation we will use in our proof.

Let $\Gamma[n] = [n]^2$ with $N(n) = |\Gamma[n]| = \binom{n}{2}$ and let $\mathcal{P} \subseteq 2^{\Gamma[n]}$ be a family of subsets of $\Gamma[n]$ for $n \in \mathbb{N}$. Let $P(n)$ be a sequence of real numbers such that $0 \leq p \leq 1$, and let $M(n)$ be a sequence of integers satisfying $0 \leq M(n) \leq N(n) = \binom{n}{2}$.

Proposition 6.20. *Let $\mathcal{P}$ be an arbitrary property of subsets of $\Gamma = \Gamma[n]$ as above, $p = p(n) \in [0,1]$ and $0 \leq a \leq 1$. If for every sequence $M = M(n)$ such that $M = \binom{n}{2}p + O(\sqrt{Np(1-p)})$ it holds that $P(G(n,M) \in \mathcal{P}) \to a$ as $n \to \infty$, then $P(G(n,p) \in \mathcal{P}) \to a$ as $n \to \infty$.*

Proof. Let C be a large constant and define for each n the set

$$\mathcal{M}(C) = \{M : |M - Np| \leq C\sqrt{Np(1-p)}\}.$$

Let X be the random variable representing the number of edges in $G(n,p)$. Then X is binomially distributed with $\mathbb{E}(X) = Np$ and $\mathrm{var}(X) = Np(1-p)$.

It follows by Chebyshev's inequality that

$$\begin{aligned} P(E(G(n,p)) \notin \mathcal{M}(C)) &= P(|X - Np| > c\sqrt{Np(1-p)}) \\ &\leq \frac{\mathrm{var}(x)}{c^2 Np(1-p)} \\ &= \frac{1}{c^2}. \end{aligned}$$

Let m_*,m^* be the elements of $\mathcal{M}(C)$ which minimize and maximize $P(G(n,m) \in \mathcal{P})$ respectively. Using the law of total probability we find,

$$\begin{aligned} P(G(n,p) \in \mathcal{P}) &= \sum_{m=1}^{N} P(E(G(n,p)) = m) \cdot P(G(n,m) \in \mathcal{P}) \\ &\geq \sum_{m \in \mathcal{M}(C)} P(E(G(n,p)) = m) \cdot P(G(n,m) \in \mathcal{P}) \\ &\geq \sum_{m \in \mathcal{M}(C)} P(E(G(n,m)) = m) \cdot P(G(n,m_*) \in \mathcal{P}). \end{aligned}$$

Hence it follows that

$$\begin{aligned} P(G(n,p) \in \mathcal{P}) &\geq P(G(n,m_*) \in \mathcal{P}) \cdot \sum_{m \in \mathcal{M}(C)} P(E(G(n,m)) = m) \\ &= P(G(n,m_* \in \mathcal{P}) \cdot P(E(G(n,p)) \in \mathcal{M}) \\ &= P(G(n,m_*) \in \mathcal{P})(1 - c^{-2}). \end{aligned}$$

Since $P(G(n,m_*) \in \mathcal{P}) = a$ as $n \to \infty$ it follows that

$$\liminf_{n \to \infty} P(G(n,p) \in \mathcal{P}) \geq a(1 - c^{-2}).$$

Now we prove an upper bound for $P(G(n,p) \in \mathcal{P})$. Again using the law of

total probability it follows that,

$$\begin{aligned} P(G(n,p) \in \mathcal{P}) &= \sum_{m=1}^{N} P(E(G(n,p)) = m) \cdot P(G(n,m) \in P) \\ &= \sum_{m \notin \mathcal{M}(C)} P(E(G(n,p)) = m) \cdot P(G(n,m) \in \mathcal{P}) \\ &+ \sum_{m \in \mathcal{M}(C)} P(E(G(n,p)) = m) \cdot P(G(n,m) \in \mathcal{P}) \\ &\leq \sum_{m \notin \mathcal{M}(C)} P(E(G(n,p)) = m) \\ &+ \sum_{m \in \mathcal{M}(C)} P(E(G(n,p)) = m) \cdot P(G(n,m) \in \mathcal{P}). \end{aligned}$$

Hence $P(G(n,p) \in \mathcal{P})$ is less than

$$P(E(G(n,p)) \notin \mathcal{M}(C)) + \sum_{m \in \mathcal{M}(C)} P(E(G(n,p) = m) P(G(n,m) \in \mathcal{P}).$$

It follows that,

$$\begin{aligned} P(G(n,p) \in \mathcal{P}) &\leq c^{-2} + \sum_{m \in \mathcal{M}(C)} P(E(G(n,p)) = m) \cdot P(G(n,m) \in \mathcal{P}) \\ &\leq c^{-2} + \sum_{m \in \mathcal{M}(C)} P(E(G(n,p)) = m)) \cdot P(G(n,m^*) \in \mathcal{P}) \\ &\leq c^{-2} + P(G(n,m^*) \in \mathcal{P}). \end{aligned}$$

Finally since $P(G(n,m^*) \in \mathcal{P}) = a$ it follows that

$$\liminf_{n \to \infty} P(G(n,p) \in P) \leq a + c^{-2}$$

and so taking c to be a large enough constant and using the upper and lower bounds that we obtained it follows that $\lim_{n \to \infty} P(G(n,p) \in P) = a$. $\square$

Chapter 7

Threshold functions

We proved in the previous section that $G(n,p) \in \mathcal{P}$ a.a.s for various monotone graph properties. In section 4.6 we saw various monotone properties which held for almost all graphs provided the value of p considered was constant. However for most properties the critical order of magnitude of p around which the property will 'just' occur or not occur will be a function of n tending to zero as $n \to \infty$. Given a monontone graph property $\mathcal{P}$, we can verify that $P(G(n,0) \in \mathcal{P}) = 0$ and $P(G(n,1) \in \mathcal{P}) = 1$. As p increases from 0 to 1, our random graph $G(n,p)$ will evolve and acquire more edges, we will want to study the nature of $P(G(n,p) \in \mathcal{P})$ (which will be a function of n) especially how quickly it approaches 1.

Definition 7.1. We call a real function $t(n)$ with $t(n) \neq 0$ for all n, a threshold function for a graph property $\mathcal{P}$ if the following holds:

$$\lim_{n\to\infty} P(G(n,p) \in \mathcal{P}) = \begin{cases} 0, \text{ if } \frac{p}{t} \to 0 \text{ as } n \to \infty \\ 1, \text{ if } \frac{p}{t} \to \infty \text{ as } n \to \infty \, . \end{cases}$$

Bollobás and Thomason proved the following important theorem on threshold functions for monotone graph properties. We will not prove it here, but we will look at many important consequences of the result.

Theorem 7.2 (**Bollobás and Thomasson** [4])**.** *Every monotone graph property has a threshold.*

When trying to compute a threshold function for a property $\mathcal{P}$, it's useful to view $\mathcal{P}$ in the following sense

$$\mathcal{P} = \{G = ([n], E) : X(G) \geq 1\}, \tag{7.1}$$

where X is a suitable non-negative random variable defined on $G(n,p)$. If for example our property was the class of graphs containing a cycle, we may have $X(G)$ count the number of cycles in G. In most cases our random variable X will be clear from context and will often be the sum of indicator random variables.
So how do we go about proving that a function t is a threshold function for a property $\mathcal{P}$? It is clear from Definition 7.1 that we must prove two things. Firstly when p is small compared to t we must show that a.a.s $G(n,p) \notin \mathcal{P}$ and secondly when p is large compared to t that a.a.s $G(n,p) \in \mathcal{P}$. Equivalently-by considering property $\mathcal{P}$ in the form of equation (6.1), we can prove t is a threshold function by proving that if p is small compared to t then a.a.s $X = 0$ and if p is large compared to t then a.a.s $X > 0$.
Since $X \geq 0$ we can use Markov's inequality for the first part of the proof. If $\mathbb{E}(X) = o(1)$ then $P(X > 0) = o(1)$ by Markov's inequality and so a.a.s $X = 0$. For the second part we want to show that a.a.s $X > 0$. Recall using corollary 4.23 that if $\frac{\text{var}(X)}{\mathbb{E}(X)^2} = o(1)$ then $P(X = 0) = o(1)$ and so $X > 0$ a.a.s. Thus it suffices prove that $\frac{\text{var}(X)}{\mathbb{E}(X)^2} = o(1)$.

We start our journey by considering a fairly simple property $\mathcal{E}$ which is the property that a graph has at least 1 edge. While it is certainly not the most interesting graph property, it is useful to illustrate the techniques used in proving that a function t is a threshold function. For all of the properties considered in this section we will simply state the threshold function and prove it , however in this example we will how it is possible to obtain a threshold function first and then prove it.

Example 7.3. What is the threshold function for the property $\mathcal{E}$?

Firstly we wish to express our property $\mathcal{E}$ in the form of equation (6.1), thus

$$\mathcal{E} = \{G = ([n], E) : X(G) > 0\}$$

where $X(G) = |E|$. If we arbitrarily enumerate all edges of K_n as $e_1, \ldots, e_N$, then we can express the random variable X as the sum of $N = \binom{n}{2}$ independent indicator random variables as follows. Let $X = \sum_{i=1}^{N} X_i$ where

$$X_i = \begin{cases} 1, \text{ if } e_i \text{ is an edge of } G(n,p) \\ 0, \text{ otherwise .} \end{cases}$$

Central to our outlined proof above is $\mathbb{E}(X)$. It is used in both cases of the outlined proof when bounding $P(X > 0)$. Thus we will obtain the threshold function of $\mathcal{E}$ by calculating when $\mathbb{E}(X) \to 0$ and $\mathbb{E}(X) \to \infty$ as $n \to \infty$.

The expectation is given as follows

$$\begin{aligned} \mathbb{E}(X) &= \sum_{i=1}^{N} \mathbb{E}(X_i) \\ &= \sum_{i=1}^{N} p \\ &= Np \\ &= \binom{n}{2} p \\ &= \Theta(n^2 p) \end{aligned}$$

the first equality follows from the linearity of expectation. It follows that if $p \ll n^{-2}$ then $\mathbb{E}(X) \to 0$ as $n \to \infty$ on the other hand if $n^{-2} \ll p$ then $\mathbb{E}(X) \to \infty$ as $n \to \infty$. Thus our candidate threshold function for $\mathcal{E}$ is $t(n) = n^{-2}$.

Claim 7.4. *$t(n) = n^{-2}$ is a threshold function for $\mathcal{E}$.*

Proof. Following from the discussion above, express $\mathcal{E}$ as

$$\mathcal{E} = \{G = ([n], E) : X(G) > 0\}$$

where $X(G) = |E|$. Arbitrarily enumerating the edges of K_n as $e_1, \ldots, e_N$, express X as the sum of n independent with $X = \sum_{i=1}^{N} X_i$ where

$$X_i = \begin{cases} 1, \text{ if } e_i \text{ is an edge of } G(n,p) \\ 0, \text{ otherwise.} \end{cases}$$

The expectation of X is given by

$$\begin{aligned}\mathbb{E}(X) &= \sum_{i=1}^{N} \mathbb{E}(X_i) \\ &= \sum_{i=1}^{N} p \\ &= \binom{n}{2} p \\ &= \Theta(n^2 p),\end{aligned}$$

with the first equality following from the linearity of expectation.
We now consider the two cases presented in Definition 7.1.

1. In the first case, consider $0 \leq p = p(n) \leq 1$ with $\frac{p}{t} \to 0$ as $n \to \infty$. Since $\frac{p}{t} = n^2 p$ and $\mathbb{E}(X) = \Theta(n^2 p)$ we conclude that $\mathbb{E}(X) \to 0$ as $n \to \infty$. Using Markov's inequality we obtain that

$$P(X > 0) \leq \mathbb{E}(X).$$

Thus $P(X > 0) \to 0$ as $n \to \infty$, hence $X = 0$ a.a.s and consequently $G(n,p) \notin \mathcal{E}$ a.a.s .

2. Now consider the second case, where $0 \leq p \leq 1$ with $\frac{p}{t} \to \infty$ as $n \to \infty$. Since $\frac{p}{t} = n^2 p$ and $\mathbb{E}(X) = \Theta(n^2 p)$ we conclude that $\mathbb{E}(X) \to \infty$ as $n \to \infty$. Next we compute

$$\begin{aligned}\frac{\mathrm{var}(X)}{\mathbb{E}(X)^2} &= \frac{Np(1-p)}{(Np)^2)} \\ &= \frac{1-p}{Np} \\ &\leq \frac{1}{Np} \\ &= \Theta((n^2 p)^{-1}),\end{aligned}$$

since $n^2 p \to \infty$ as $n \to \infty$ we conclude that $\frac{\mathrm{var}(X)}{\mathbb{E}(X)^2} = o(1)$. Finally it follows from corollary 4.23 that $P(X = 0) = o(1)$. Thus $P(X > 0) \to 1$ as $n \to \infty$, hence $X \geq 1$ a.a.s and consequently $G(n,p) \in \mathcal{E}$ a.a.s . □

After considering perhaps the simplest graph property $\mathcal{E}$, we will now focus our attention on more interesting properties. As we increase the number of edges in our random graph (by increasing p) it will obtain more and more structure. As more edges are introduced, our random graph will start forming trees. Thus the next property we will consider will be the existence of a tree on r vertices.

Theorem 7.5. *Let $2 < r \leq n$ and let $\mathcal{T}_r$ be the property of containing a tree on r vertices. Then $t(n) = n^{-\frac{r}{r-1}}$ is a threshold for $\mathcal{T}_r$.*

Proof. Cayley's formula tells us that in a complete graph on r labeled vertices there are precisely r^{r-2} labeled spanning trees [6]. We begin by enumerating all possible trees on r vertices that $G(n,p)$ can contain. Start by arbitrarily enumerating all sets of r vertices from $1, \ldots, \binom{n}{r}$. Next for each set of r vertices $1 \leq i \leq \binom{n}{r}$, arbitrarily enumerate all possible trees from $1, \ldots, r^{r-2}$. Denote tree $1 \leq j \leq r^{r-2}$ on vertex set i as T_{ij} and define X_{ij} to be the indicator random variable such

$$X_{ij} = \begin{cases} 1, \text{ if } T_{ij} \text{ is a tree in } G(n,p) \\ 0, \text{ otherwise .} \end{cases}$$

It follows that $P(X_{ij} = 1) = p^{r-1}$ for all $1 \leq i \leq \binom{n}{r}$ and $1 \leq j \leq r^{r-2}$ as all $r-1$ edges of the tree must be present.
The random variable

$$X = \sum_{i=1}^{\binom{n}{r}} \sum_{j=1}^{r^{r-2}} X_{ij}$$

counts the number of spanning trees of size r in $G(n,p)$. Next we compute

the expectation of the random variable X, it follows that

$$\begin{aligned}\mathbb{E}(X) &= \sum_{i=1}^{\binom{n}{r}} \sum_{j=1}^{r^{r-2}} \mathbb{E}(X_{ij}) \\ &= \sum_{i=1}^{\binom{n}{r}} \sum_{j=1}^{r^{r-2}} p^{r-1} \\ &= \binom{n}{r} r^{r-2} p^{r-1} \\ &= \Theta(r^{r-2}(n^{\frac{r}{r-1}}p)^{r-1}).\end{aligned}$$

We now consider the two cases presented in Definition 7.1

1. In the first case consider $0 \leq p \leq 1$ with $\frac{p}{t} \to 0$ as $n \to 0$. Since $\frac{p}{t} = n^{\frac{r}{r-1}}p \to 0$ as $n \to \infty$ it follows that $(n^{\frac{r}{r-1}}p)^{r-1} \to 0$ as $n \to \infty$. Furthermore as $\mathbb{E}(X) = \Theta(r^{r-2}(n^{\frac{r}{r-1}}p)^{r-1})$ we conclude that $\mathbb{E}(X) \to 0$ as $n \to \infty$. Using Markov's inequality we obtain

   $$P(X > 0) \leq \mathbb{E}(X).$$

   Thus $P(X > 0) \to 0$ as $n \to 0$, hence $X = 0$ a.a.s and consequently $G(n,p) \notin \mathcal{T}_r$ a.a.s.

2. Now consider the second case, where $0 \leq p \leq 1$ with $\frac{p}{t} \to \infty$ as $n \to \infty$. Furthermore as $\mathbb{E}(X) = \Theta(r^{r-2}(n^{\frac{r}{r-1}}p)^{r-1})$ we conclude that $\mathbb{E}(X) \to 0$ as $n \to \infty$. Next we compute

   $$\begin{aligned}\frac{\mathrm{var}(X)}{\mathbb{E}(X)^2} &= \frac{\binom{n}{r} r^{r-2} p^r (1-p^r)}{(\binom{n}{r} r^{r-2} p^r)^2} \\ &\leq \frac{1}{r^{r-2}\binom{n}{r} p^{r-1}} \\ &= \frac{1}{\mathbb{E}(X)},\end{aligned}$$

   since $\mathbb{E}(X) \to \infty$ as $n \to \infty$ we conclude that $\frac{\mathrm{var}(X)}{\mathbb{E}(X)^2} = o(1)$. Finally it follows from corollary 4.23 that $P(X = 0) = o(1)$. Thus $P(X > 0) \to 1$ as $n \to \infty$, hence $X \geq 1$ a.a.s and consequently $G(n,p) \in \mathcal{T}_r$ a.a.s . □

We now move on to the study of clique numbers. Recall that the clique number of a graph is the size of it's largest clique. Naturally as our graph obtains more edges we expect the clique number to be higher. Let $\mathcal{C}_{l4}$ be the property of having a clique number of at least 4. We will investigate the threshold function for $\mathcal{C}_{l4}$.

Theorem 7.6. *The property $\mathcal{C}_{l4}$ has threshold function $t(n) = n^{-2/3}$.*

Proof. We start by expressing the property $\mathcal{C}_{l4}$ in the form of equation 6.1. Given a random graph G let $X(G)$ be the random variable that counts the number of cliques of size four. Then clearly $\omega(G) \geq 4$ if and only $X(G) \geq 1$. Hence

$$\mathcal{C}_{l4} = \{G = ([n], E) : X(G) \geq 1\}.$$

We can express the random variable X as the sum of indicator random variables in the following way. Start by arbitrarily enumerating all sets of vertices of size 4 from S_1 to $S_{\binom{n}{4}}$. For set S_i with $1 \leq i \leq \binom{n}{4}$ let T_i be the event that S_i forms a clique and define X_i to be the indicator random variable such that

$$X_i = \begin{cases} 1, \text{ if } G(n,p) \text{ contains a clique on vertex set } S_i \\ 0, \text{ otherwise .} \end{cases}$$

We can now express the random variable X as

$$X = \sum_{i=1}^{\binom{n}{4}} X_i.$$

The expectation of the random variable X is then

$$\begin{aligned} \mathbb{E}(X) &= \sum_{i=1}^{\binom{n}{4}} \mathbb{E}(X_i) \\ &= \sum_{i=1}^{\binom{n}{4}} p^6 \\ &= \binom{n}{4} p^6 \\ &= \Theta(n^4 p^6). \end{aligned}$$

We now consider the two cases presented in Definition 7.1.

1. In the first case consider $0 \leq p \leq 1$ with $\frac{p}{t} \to 0$ as $n \to 0$. Since $\frac{p}{t} = pn^{\frac{2}{3}}$ and $pn^{\frac{2}{3}} \to 0$ as $n \to 0$ it follows that $(pn^{\frac{2}{3}})^6 = p^6n^4 \to 0$ as $n \to 0$. Furthermore as $\mathbb{E}(X) = \Theta(n^4p^6)$ we conclude that $\mathbb{E}(X) \to 0$ as $n \to 0$. Using Markov's inequality we obtain that

$$P(X > 0) \quad \leq \quad \mathbb{E}(X).$$

Thus $P(X > 0) \to 0$ as $n \to \infty$, hence $X = 0$ a.a.s and consequently $G(n,p) \notin \mathcal{C}_{l4}$ a.a.s .

2. Now consider the second case $0 \leq p \leq 1$ with $\frac{p}{t} \to \infty$ as $n \to \infty$. Since $\frac{p}{t} = pn^{\frac{2}{3}}$ we conclude $n^4p^6 \to \infty$ as $n \to \infty$ and thus $\mathbb{E}(X) \to \infty$ as $n \to \infty$. We now consider the Δ^* of Corollary 4.29 (since all 4-sets look the same and so X_i are symmetric). For any two sets of 4 vertices S_i and S_j, $i \sim j$ if and only if $i \neq j$ and S_i and S_j have common edges i.e. $|S_i \cap S_j| = 2$ or 3. For fixed S_i there are $\binom{4}{2}\binom{n-4}{2} = O(n^2)$ sets S_j such that $|S_i \cap S_j| = 2$ with $P(T_j|T_i) = p^5$, furthermore there are $\binom{4}{3}\binom{n-4}{1} = O(n)$ sets S_k such that $|S_k \cap S_i| = 3$ with $P(T_k|T_i) = p^3$. Thus

$$\begin{aligned} \Delta^* &= \sum_{j \sim i} P(T_j|T_i) \\ &= O(n^2p^5) + O(np^3) \\ &= o(n^4p^6) \\ &= o(\mathbb{E}(X)) \end{aligned}$$

with the third equality following from the fact that $pn^{\frac{2}{3}} \to \infty$ as $n \to \infty$. Thus using Corollary 4.29 it follows that $X > 0$ a.a.s so $G(n,p) \in \mathcal{C}_{l4}$ a.a.s. □

7.1 Balanced Graphs

All the graph properties we have considered so far in this section have been classes of graphs containing some subgraph H. Furthermore all threshold functions for these properties seem to be following a similar pattern, namely if H is our subgraph, then threshold function for H is $t(n) = n^{-\frac{v}{e}}$. In this

section we will prove a result which generalizes the ones we have seen in the previous section.

Definition 7.7. Let H be a graph with $v \geq 1$ vertices and e edges. We call $\rho(H) = \frac{e}{v}$ the density of H.

Definition 7.8. We call a graph H balanced if every subgraph H' satisfies $\rho(H') \leq \rho(H)$. We call H strictly balanced if every proper subgraph H' satisfies $\rho(H') < \rho(H)$.

Example 7.9. We start by showing that K_n is a strictly balanced graph. Firstly $\rho(K_n) = \frac{\binom{n}{2}}{n} = \frac{n-1}{2}$, thus any proper subgraph $G' \subset K_n$ on n vertices will have less edges than K_n, thus $\rho(G') < \rho(K_n)$. Furthermore any proper subgraph $G'' \subset K_n$ on $k < n$ vertices has maximum density

$$\rho(G'') \leq \frac{\binom{k}{2}}{k} = \frac{k-1}{2} < \frac{n-1}{2}.$$

On the other hand consider a labeled graph G on $n+1$ vertices consisting of K_n with vertices $1, \ldots, n$ and an edge from vertex n to vertex $n+1$. Then

$$\rho(G) = \frac{\binom{n}{2} + 1}{n+1} < \frac{\binom{n}{2}}{n}$$

and so G contains a subgraph of higher density-namely K_n, and so is not balanced.

Theorem 7.10. *Let H be a balanced graph with v vertices and $e \geq 1$ edges. Let $\mathcal{P}_H$ be the property of containing H as a subgraph. Then $t(n) = n^{-\frac{v(H)}{e(H)}}$ is a threshold function for $\mathcal{P}_H$.*

Proof. We start by expressing the property $\mathcal{P}_H$ in the form of equation 6.1. Given a random graph G let $X(G)$ be the random variable that counts the number of copies of H in G. We can express the random variable X as the sum of indicator random variables in the following way. Start by arbitrarily enumerating all sets of vertices of size v as $S_1, \ldots, S_{\binom{n}{v}}$, and let T_i be the

event that $G(n,p)$ contains a copy of H on vertex set S_i. Define X_i to be the indicator random variable such that

$$X_i = \begin{cases} 1, \text{ if } G(n,p) \text{ contains } H \text{ on vertex set } S_i \\ 0, \text{ otherwise .} \end{cases}$$

We can express the random variable X as

$$X = \sum_{i=1}^{\binom{n}{v}} X_i.$$

We bound $P(X_i = 1)$ as follows

$$p^e \leq P(X_i = 1) \leq v! p^e \tag{7.2}$$

The lower bound follows because the probability of the e edges required to form H is p^e, the upper bound arises by considering all possible labeling's of the vertices of S_i.
The expectation of the random variable X is given by

$$\begin{aligned} \mathbb{E}(X) &= \sum_{i=1}^{\binom{n}{v}} \mathbb{E}(X_i) \\ &= \sum_{i=1}^{\binom{n}{v}} P(X_i = 1) \\ &= \binom{n}{4} P(X_1 = 1) \\ &= \Theta(n^v p^e). \end{aligned}$$

The first equality follows from the linearity of expectation, the third equality follows since $P(X_i = 1)$ is the same for all $1 \leq i \leq \binom{n}{4}$ and the final equality follows from 6.2.

We now consider the two cases presented in Definition 7.1.

1. In the first case consider $0 \leq p \leq 1$ with $\frac{p}{t} \to 0$ as $n \to 0$. Since $\frac{p}{t} = pn^{\frac{v}{e}}$ and $pn^{\frac{v}{e}} \to 0$ and $n \to 0$ it follows that $(pn^{\frac{v}{e}})^e = p^e n^v \to 0$ as $n \to 0$.

Furthermore as $\mathbb{E}(X) = \Theta(n^v p^e)$ we conclude that $\mathbb{E}(X) \to 0$ as $n \to 0$. Using Markov's inequality we obtain that

$$P(X > 0) \quad \leq \quad \mathbb{E}(X).$$

Thus $P(X > 0) \to 0$ as $n \to \infty$, hence $X = 0$ a.a.s and consequently $G(n,p) \notin \mathcal{P}_H$ a.a.s .

2. Now consider the second case $0 \leq p \leq 1$ with $\frac{p}{t} \to \infty$ as $n \to \infty$. Since $\frac{p}{t} = pn^{\frac{v}{e}}$ we conclude $n^v p^e \to \infty$ as $n \to \infty$ and thus $\mathbb{E}(X) \to \infty$ as $n \to \infty$. We now consider the Δ^* of Corollary 4.29 (since all v-sets look the same and so X_i are symmetric). As in the previous case, $i \sim j$ if $S_i \neq S_j$ and $|S_i \cap S_j| = i$, for $2 \leq i \leq v-1$. It will be useful to express Δ^* in the following way

$$\begin{aligned} \Delta^* &= \sum_{i \sim j} P(T_j | T_i) \\ &= \sum_{i=2}^{v-1} \sum_{|i \cap j| = i} P(T_j | T_i). \end{aligned}$$

For a fixed set S_i and for each $2 \leq k \leq v-1$ there are $\binom{v}{i}\binom{n-v}{n-i} = O(n^{v-i})$ choices of S_j such that $|S_i \cap S_j| = k$. Given fixed S_i and S_j we wish to calculate $P(T_j|T_i)$. Since H is balanced the subgraph formed on the k vertices which belong to both S_i and S_j cannot have more than $\frac{ke}{v}$ edges, thus if S_j contains a copy of H at least $e - \frac{ke}{v}$ of these edges have at least one vertex outside S_i. Hence

$$P(T_j | T_i) = O(p^{e - \frac{ke}{v}})$$

and

$$\begin{aligned} \Delta^* &= \sum_{k=2}^{v-1} O(n^{v-k} p^{e - \frac{ke}{v}}) \\ &= \sum_{k=2}^{v-1} O((n^v p^e)^{1 - \frac{k}{v}}) \\ &= \sum_{k=2}^{v-1} o(n^v p^e) \\ &= o(\mathbb{E}(X)). \end{aligned}$$

The third equality following from the fact that $n^v p^e \to \infty$ as $n \to \infty$. Thus using Corollary 4.29 it follows that $X > 0$ a.a.s so $G(n,p) \in \mathcal{P}_H$ a.a.s. □

Let us now look at the case where H is not balanced, does the result hold in the unbalanced case too? The answer is no, and we prove it below.

Claim 7.11. *If H is a subgraph with v vertices and e edges that is not balanced and $\mathcal{P}$ is the property that $G(n,p)$ contains H Then $t(n) = p^{-\frac{v}{e}}$ is not a threshold function for $\mathcal{P}$.*

Proof. Suppose not, let H_1 be the subgraph with v_1 vertices and e_1 edges and $\frac{e_1}{v_1} > \frac{e}{v}$. Let α satisfy $\frac{v}{e} < \alpha < \frac{v_1}{e_1}$ and let $p(n) = n^{-\alpha}$. Then $G(n,p)$ will almost surely contain a copy of H but no copy of H_1. However $H_1 \subseteq H$ and so this is impossible contradicting that $p^{-\frac{v}{e}}$ is a threshold for H. □

7.2 Sharp Thresholds

The behavior of threshold functions exhibit very typical behaviors at extreme values of p (i.e. near 0 or 1). For a given monotone graph property $\mathcal{P}$ the function $P(G(n,p) \in \mathcal{P})$ will be very close to 0 when p is near 0 and will be close to 1 when p is close to 1. Theorem 7.2 allows us to be more specific about the range of values of p where $P(G(n,p) \in \mathcal{P})$ is close to 0 or 1. If $t(n)$ is a threshold function for $\mathcal{P}$ it follows that when $p \ll t(n)$ our property will not hold with high probability and when $p \gg t(n)$ our property will hold with high probability. It is often a lot less clear how $P(G(n,p) \in \mathcal{P})$ behaves when p to be close to the value $t(n)$. For certain monotone graph properties the function $P(G(n,p) \in \mathcal{P})$ increases sharply either side of $t(n)$ from 0 to 1. Informally if $t(n)$ is a threshold function for the property $\mathcal{P}$ which portrays this behavior we say that it is a sharp threshold.

Definition 7.12. The function t is a sharp threshold for monotone graph property $\mathcal{P}$ if for all $\epsilon > 0$

$$\lim_{n\to\infty} P(G(n,p) \in \mathcal{P}) = \begin{cases} 0, \text{ if } p < (1-\epsilon)t \\ 1, \text{ if } p > (1+\epsilon)t. \end{cases}$$

7.2.1 Sharp Threshold for Connectivity

Connectivity is one of the most fundamental properties in Graph theory let us denote the property of being connceted by $\mathcal{C}$. We know from Theorem 7.2 that $\mathcal{C}$ has a threshold but it turns out that the threshold is sharp. What is more interesting, is that for values of p below this threshold value we are almost certain to have isolated vertices. Thus as we increase p from 0 to 1, the moment the last isolated vertex in our random graph disappears our graph becomes connected.

Theorem 7.13 (Erdős-Réyni)**.** *A sharp theshold for the connectivity for* $G(n,p)$ *is* $t(n) = \frac{\log(n)}{n}$.

Proof. We must prove both of the following claims;

1. If $0 < \epsilon < 1$ and $p = (1-\epsilon)\frac{\log(n)}{n}$ then $P(G(n,p) \text{ is connected}) \to 0$ as $n \to \infty$
2. If $0 < \epsilon$ and $p = (1+\epsilon)\frac{\log(n)}{n}$ then $P(G(n,p) \text{ is connected}) \to 1$ as $n \to \infty$.

We start by proving 1

1. Since a graph with an isolated vertex is not connected we start by counting the number of isolated vertices in $G(n,p)$. Start be arbitrarily enumerating the vertices from 1 to n. For vertex $1 \leq i \leq n$ let X_i be an indicator random variable defined as follows

$$X_i = \begin{cases} 1, \text{ if node is isolated} \\ 0, \text{ otherwise .} \end{cases}$$

We are interested in the expected number of isolated vertices in order to calculate the expectation we must first calculate $P(X_i = 1)$. It follows that

$$\begin{aligned} P(X_i = 1) &= (1-p)^{n-1} \\ &= \Theta(e^{-p(n-1)}) \\ &= \Theta(n^{-(1-\epsilon)\frac{n-1}{n}}) \\ &= \Theta(n^{-(1-\epsilon)}). \end{aligned}$$

The third inequality follows since $e^{-p(n-1)} = n^{-(1-\epsilon)\frac{n-1}{n}}$ by substituting $p = (1-\epsilon)\frac{\log(n)}{n}$. Define the random variable X to be the total number

of isolated vertices. Hence

$$X = \sum_{i=1}^{n} X_i.$$

and the expectation of X is given by

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{i=1}^{n} \mathbb{E}(X_i) \\
&= \sum_{i=1}^{n} P(X_i = 1) \\
&= nP(X_1 = 1) \\
&= n\Theta(n^{-(1-\epsilon)}) \\
&= \Theta(n^{1-(1-\epsilon)}) \\
&= \Theta(n^{\epsilon}).
\end{aligned}$$

The first equality follows from the linearity of expectation and the third inequality holds since $P(X_i = 1)$ is the same for all $1 \leq i \leq n$. For $0 < \epsilon < 1$, we have $\mathbb{E}(X) \to \infty$. It follows from Corollary 4.23 that if $\frac{\text{var}(X)}{\mathbb{E}(X)^2} = o(1)$ the $X > 0$ a.a.s and thus we have at least one isolated vertex. Hence $G(n,p)$ is not connected a.a.s .

Next we compute $\text{var}(X)$. Using Theorem 3.39 and denoting

$$q = P(X_i = 1) = (1-p)^{n-1}$$

for $1 \leq i \leq n$ we obtain the following,

$$\begin{aligned}
\text{var}(X) &= \sum_{i=1}^{n} \text{var}(X_i) + 2\sum_{i<j} \text{cov}(X_i, X_j) \\
&= n \cdot \text{var}(X_1) + 2\frac{n(n-1)}{2} \cdot \text{cov}(X_1, X_2) \\
&= nq(1-q) + n(n-1)(\mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)) \\
&= nq(1-q) + n(n-1)(\mathbb{E}(X_1X_2) - q^2).
\end{aligned}$$

The second inequality follows since $\text{var}(X_i)$ and $\text{cov}(X_i, X_j)$ are the same for all $1 \leq i \leq j$ and $1 \leq i < j \leq n$ respectively.

Thus to compute var(X) we must compute $\mathbb{E}(X_1X_2)$. The random variable $X_1X_2 = 1$ if and only if both vertex 1 and 2 are isolated i.e $X_1 = 1$ and $X_2 = 1$. If follows then that the edge between vertex 1 and 2 cannot be present and neither can the $2(n-2)$ edges from vertices 1 and 2 to the remaining $n-2$ vertices in $G(n,p)$. Thus for vertices 1 and 2 to be isolated the $2(n-2)+1 = 2n-3$ edges mentioned cannot be present. Hence $P(X_1X_2 = 1) = (1-p)^{2n-3}$ and the expected value of the random variable X_1X_2 is given by

$$\begin{aligned}\mathbb{E}(X_1X_2) &= P(X_1X_2 = 1)\\ &= (1-p)^{2n-3}\\ &= \frac{q^2}{(1-p)}.\end{aligned}$$

It follows that

$$\begin{aligned}\text{var}(X) &= nq(1-q) + n(n-1)\left(\frac{q^2}{(1-p)} - q^2\right)\\ &= nq(1-q) + n(n-1)\frac{q^2p}{(1-p)}\\ &< nq + 2n^2q^2p\\ &< 2(nq + n^2q^2p).\end{aligned}$$

Recall $q = \Theta(n^{-(1-\epsilon)})$ and $p = (1-\epsilon)\frac{\log(n)}{n}$ and so

$$\begin{aligned}nq + n^2q^2p &= O(nn^{-(1-\epsilon)} + n(1-\epsilon)\log(n)n^{-2(1-\epsilon)})\\ &= O(2nn^{-(1-\epsilon)})\\ &= 2O(nn^{-(1-\epsilon)})\\ &= O(\mathbb{E}(X)).\end{aligned}$$

The second equality follows since $(1-\epsilon)\log(n)n^{-2(1-\epsilon)} = O(n^{-(1-\epsilon)})$.

Thus var$(X) = O(\mathbb{E}(X))$ and since $\mathbb{E}(X) \to \infty$ as $n \to \infty$ it follows that var$(X) = O(\mathbb{E}(X)^2)$. Using Corollary 4.23 we conclude that $X > 0$ a.a.s and hence $G(n,p) \notin \mathcal{C}$ a.a.s.

2. Claim 2 is equivalent to the statement that, if $p = (1+\epsilon)\frac{\log(n)}{n}$ then
$$P(G(n,p) \text{ is disconnected}) \to 0 \text{ as } n \to \infty.$$

Any disconnected graph on n vertices contains a vertex cut, i.e the vertex $[n]$ set can be partitioned into two sets of size k and $n-k$ where $1 \leq k \leq n/2$ with no edges between them. If we partition $[n]$ into two sets A_k and $B_k = [n] \setminus A_k$ with $|X_k| = k$, then there are exactly $k(n-k)$ potential edges between the two sets. Thus the probability that there are no edges between the sets X_k and Y_k is $(1-p)^{k(n-k)}$. Let X be the random variable which counts the number of vertex cuts in $G(n,p)$. Then $G(n,p)$ is disconnected if $X > 0$, and so we wish to show if $p = (1+\epsilon)\frac{\log(n)}{n}$ with $\epsilon > 0$ then $X > 0$ a.a.s . Start by arbitrarily enumerating all sets of k vertices from 1 to $\binom{n}{k}$ for $1 \leq k \leq \frac{n}{2}$. Let A_{jk} refer to set $1 \leq j \leq \binom{n}{k}$ of size $1 \leq k \leq \frac{n}{2}$ and let $B_{jk} = [n] \setminus A_{jk}$. Let X_{jk} be the indicator random variable such that

$$X_{jk} = \begin{cases} 1, \text{ if there are no edges between } A_{jk} \text{ and } B_{jk} \\ 0, \text{ otherwise }, \end{cases}$$

and it follows that $P(X_{jk} = 1) = (1-p)^{k(n-k)}$. The random variable $X = \sum_{k=1}^{\frac{n}{2}} \sum_{j=1}^{\binom{n}{k}} X_{jk}$ counts the number of vertex cuts in $G(n,p)$.
We are interested in calculating the expected number of vertex cuts in $G(n,p)$, using Markov's inequality it follows that if $\mathbb{E}(X) \to 0$ as $n \to \infty$ then $X = 0$ a.a.s and consequently $G(n,p) \in \mathcal{C}$ as $G(n,p)$ has no vertex cuts. The expectation of X is given by

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=1}^{\frac{n}{2}} \sum_{j=1}^{\binom{n}{k}} \mathbb{E}(X_{jk}) \\ &= \sum_{k=1}^{\frac{n}{2}} \sum_{j=1}^{\binom{n}{k}} P(X_{jk} = 1) \\ &= \sum_{k=1}^{\frac{n}{2}} \sum_{j=1}^{\binom{n}{k}} (1-p)^{k(n-k)} \\ &= \sum_{k=1}^{\frac{n}{2}} \binom{n}{k} (1-p)^{k(n-k)} \end{aligned}$$

with first inequality following from the linearity of expectation. Furthermore

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=1}^{\frac{n}{2}} \binom{n}{k}(1-p)^{k(n-k)} \\
&= \sum_{k=1}^{\lfloor\sqrt{n}\rfloor} \binom{n}{k}(1-p)^{k(n-k)} + \sum_{k=\lfloor\sqrt{n}\rfloor+1}^{\frac{n}{2}} \binom{n}{k}(1-p)^{k(n-k)} \\
&\leq \sum_{k=1}^{\lfloor\sqrt{n}\rfloor} n^k (e^{-p})^{k(n-k)} + \sum_{k=\lfloor\sqrt{n}\rfloor+1}^{\frac{n}{2}} \left(\frac{ne}{k}\right)^k (e^{-p})^{k(n-k)} \\
&= \sum_{k=1}^{\lfloor\sqrt{n}\rfloor} (n(e^{-p})^{(n-k)})^k + \sum_{k=\lfloor\sqrt{n}\rfloor+1}^{\frac{n}{2}} \left(\left(\frac{ne}{k}\right)(e^{-p})^{(n-k)}\right)^k \\
&\leq \sum_{k=1}^{\lfloor\sqrt{n}\rfloor} (n(e^{-p})^{(n-\sqrt{n})})^k + \sum_{k=\lfloor\sqrt{n}\rfloor+1}^{\frac{n}{2}} \left(\left(\frac{ne}{\sqrt{n}}\right)(e^{-p})^{(n-n/2)}\right)^k \\
&= \sum_{k=1}^{\lfloor\sqrt{n}\rfloor} (n^{\frac{1+\epsilon}{\sqrt{n}}-\epsilon})^k + \sum_{k=\lfloor\sqrt{n}\rfloor+1}^{\frac{n}{2}} (en^{\frac{-\epsilon}{2}})^k \\
&\leq \sum_{k=1}^{\lfloor\sqrt{n}\rfloor} (n^{\frac{-\epsilon}{2}})^k + \sum_{k=\lfloor\sqrt{n}\rfloor+1}^{\frac{n}{2}} (en^{\frac{-\epsilon}{2}})^k \\
&= o(1),
\end{aligned}$$

the fourth equality follows from setting $p = \frac{(1+\epsilon)\log(n)}{n}$.

Since $\mathbb{E}(X) \to 0$ as $n \to 0$ we conclude $X = 0$ a.a.s from Markov's inequality. Thus $G(n,p)$ contains no vertex cuts a.a.s and hence $G(n,p) \in \mathcal{C}$ a.a.s . □

Chapter 8

Conclusion

One of the aims of this thesis was to compare the uniform random graph model with the binomial random graph model using some simple examples. It was evident that proving that $G(n, 1/2)$ was connected a.a.s was a lot easier than proving $G(n, \binom{n}{2}/2)$ was connected a.a.s. It is because edges in the Binomial model are selected independently of one another that the model is much easier to work with than the uniform model. Results can be obtained with far less effort which is why the uniform model is now rarely studied directly. We showed that the two models are asymptotically equivalent when $M = \binom{n}{2}p$. This proves useful when we want to study the asymptotic nature of certain properties when the number of edges plays an important role, while still wanting to avoid difficult calculations faced when working with $G(n, m)$.

The first major proof using random graphs that we provided was for the result $R(k, k) \geq 2^{k/2}$, which was proved by Erdős in 1947 [11]. There are a number of different versions of this proof. Our proof showed that the probability that a random graph has a clique of size at least k is less than $1/2$, similarly the probability it has an independent set of size at least k is also less than $1/2$. Thus there exists a graph with clique number less than k and independence number less than k proving the result. It's worth noting while the argument seems trivial today it was far from obvious when originally published. In particular mathematician Paul Turán conjectured $R(k, k)$ was roughly k^2 and that Erdős' result was surprising, as the quantity behaved a lot differently than he expected.

The second major proof which used random graphs that we presented was

the proof of the existence of a graph with arbitrarily high girth and arbitrarily high chromatic number, also proved by Erdős [9]. We showed when defining the chromatic number of a graph that it is possible to construct triangle free graphs which have an arbitrarily large chromatic number [14]. Erdős used the following general idea: if the probability of each edge being selected is small, our random graph will have few short cycles. On the other hand if the probability of each edge existing is large then our random graph will not have a large independent set. Choosing the right value for the probability of each edge being selected ensures that we can delete a vertex from each short cycle while maintaining that the graph obtained does not have a large independent set and thus has large chromatic number. The technique employed by Erdős has since been extended in a number of ways most notably by Bollobás and Sauer [5] in the case of uniquely k-colourable graphs and k-critical graphs.

Next we introduced the notion of a threshold function for a monotone graph property. A result central in this area, due to Bollobás and Thomason [4], asserts that, all monotone graph properties have a threshold function. We first determined the threshold functions for some simple graph properties using the first and second moment methods. The monotone properties considered involved containing a specific subgraph in particular trees and cliques. Each of these structures are types of balanced graphs and the threshold functions for these balanced graphs exhibit a particular pattern. This lead us to prove a generalized formula for the threshold function of all balanced graphs. The study of threshold functions then lead us to introduce the concept of a sharp threshold function. Finally we presented and determined the sharp threshold function for the connectivity of the binomial random graph.

The theory of random graphs is vast, and in this thesis we cover only a fraction of it. In addition to the uniform random graph and binomial random graph models, many other random graph models have emerged. Two requiring particular reference are random planar graphs and random regular graphs. A random planar graph on n vertices is a graph drawn uniformly at random from the set of all labeled planar graphs on n vertices. One of the reasons why the analysis of random planar graphs is difficult is because edges do not occur independently of one another. In addition, while asymptotic estimates can be obtained, there does not exist a closed formula for the exact number of planar graphs on n vertices. The random regular graph $G_{n,r}$ is the probability space of r-regular graphs on n vertices, with each graph hav-

ing equal probability. One of the fundamental questions involving random regular graphs is, how do we generate one? One such approach proposed by Bollobás is to follow the configuration model [3]. The configuration model always yields a d-regular graph but unfortunately it does not always produce a simple graph (a graph with no multiple edges). The probability that one gets a simple graph is bounded by a positive constant when d is a constant and conditioning on this event, the sampling is uniform. For many practical and theoretical purposes it is usually sufficient to generate random d-regular graphs which are asymptotically uniform.

A relatively new area of research is the study of the resilience of random graphs. Given a monotone graph property $\mathcal{P}$ and a graph G, the resilience of G with respect to $\mathcal{P}$ (denoted $f_{\mathcal{P}}(G)$) is a measure of how strongly G contains the property $\mathcal{P}$. The value of the function $f_{\mathcal{P}}(G)$ can change significantly from graph to graph and for some graphs can be very difficult to determine. It is therefore natural to try and study the typical behavior of this function and this leads to the study of resilience of random graphs. The paper by B.Sudakov and V.H.Vu [20] provides an excellent introduction to the topic of resilience of random graphs.

The study of pseudo-random graphs is another area which has received a lot of attention following the development of Random Graph theory. How can we tell if a given graph behaves like a random one? Informally a pseudo-random graph is a graph $G = (V, E)$ that behaves like a random graph with the same edge density $G = (|V|, p)$ where $p = |E|/\binom{|V|}{2}$. What we mean by 'behaves like' depends on the definition of pseudo-random you are working with. It was Mathematician Andrew Thomason who launched a systematic study of pseudo-random graphs in the mid-eighties. He introduced the notion of jumbled graphs, providing us with a quantitative measure for the similarity between the edge distributions of pseudo-random and truly random graphs. The survey by M.Krivelevich and B.Sudakov [13] provides a systematic coverage of the current research in pseudo-random graphs.

Bibliography

[1] N.Alon , J.Spencer. *The Probabilistic Method*, John Wiley and Sons (2011).

[2] B.Bollobás. *Modern Graph Theory*, Springer (1998).

[3] B.Bollobás. *Random Graphs*, Cambridge University Press (2001).

[4] B.Bollobás, A.Thomason. *Threshold functions*, Combinatorica 7, no. 1, 3538 (1987)

[5] B. Bollobs, N. Sauer, *Uniquely colourable graphs with large girth*, Canad. J. Math., 28, pp. 13401344 (1970)

[6] A.Cayley. *A theorem on trees*, Quart. J. Math 23: 376378 (1889).

[7] K.L.Chung. *A Course in Probability Theory*, Academic Press (2001).

[8] R.Diestel. *Graph Theory*, Springer (2012).

[9] P.Erdős. *Graph Theory and Probability*, Canad. J. Math 11, 34-38 (1959).

[10] P.Erdős, A.P.Rényi. *On Random Graphs. I* Publicationes Mathematicae 6: 290297 (1959)

[11] P.Erdős. *Some remarks on the theory of graphs*, Bull.Am.Math.Soc 53:292-294 (1947).

[12] E.N.Gilbert, *Random graphs*, Annals of Mathematical Statistics (1959)

[13] S.Janson. Łuczak T, Ruciński A *Random Graphs*, John Wiley and Sons (2011).

[14] M.Krivelavich, B.Sudakov. *Pseudo Random Graphs*, More Sets, Graphs and Numbers Bolyai Society Mathematical Studies Volume 15 , pp 199-262 (2006)

[15] J. Mycielski. *Sur le coloriage des graphs*, Colloq. Math., 3:161162, (1955).

[16] M.E.J.Newman. *Networks: An Introduction*, Oxford University Press (2010).

[17] M.E.J.Newman. *The Structure and Dynamics of Networks*, Princeton University Press (2006).

[18] F.P. Ramsey. *On a problem in formal logic*, Proc. Lond. Math. Soc., 30:264286, (1930).

[19] V.K.Rohatgi, A.K.M.E.Saleh. *An Introduction to Probability and Statistics*, John Wiley and Sons (2011).

[20] J Spencer. *Ten Lectures on the Probabilistic Method*, SIAM (1994).

[21] B.Sudakov, V.Vu. *Local resilience of graphs*, Random Structures Algorithms Volume 33, Issue 4, pages 409433, December (2008)

[22] D. B. West. *Introduction to Graph Theory*, Prentice Hall, 2^{nd} edition, (2001).