THE GIANT COMPONENT: RANDOM GRAPHS AND INFORMATION CASCADES

Daniel Thornton

Supervisors: Prof. Catherine Greenhill, Prof. Richard Holden

School of Mathematics and Statistics
UNSW Sydney

November 2019

Submitted in partial fulfillment of the requirements of the degree of
Bachelor of Science with Honours

Plagiarism statement

I declare that this thesis is my own work, except where acknowledged, and has not been submitted for academic credit elsewhere.

I acknowledge that the assessor of this thesis may, for the purpose of assessing it:

- Reproduce it and provide a copy to another member of the University; and/or,
- Communicate a copy of it to a plagiarism checking service (which may then retain a copy of it on its database for the purpose of future plagiarism checking).

I certify that I have read and understood the University Rules in respect of Student Academic Misconduct, and am aware of any potential plagiarism penalties which may apply.

By signing this declaration I am agreeing to the statements and conditions above.

Signed: _____ Date: _____

Acknowledgements

To my darling wife Chelsea. Thank you for supporting me through the last 4 years, and especially this year. I could not have completed this thesis without all of the days you spent looking after our daughter, all of the coffee dates you took me on to clear my head, and all of your listening to my incomprehensible rambling. Thank you for forcing me to explain my thesis to a "normal" person (as you would say). Every day that I have spent writing this thesis is a day that I think about the next season of life that we are now embarking upon together, as we end this journey and begin a new one, and with a new daughter too. You are a wonderful woman, and you deserve more praise than I can give you in this short space, I love you very much.

To my supervisors Catherine Greenhill and Richard Holden. Catherine, thank you for the endless meetings and for your constant invaluable feedback. Most of all, thank you for getting me interested in this project in the first place. I had no idea what random graphs were this time last year, and we have come so far. I am really grateful that once we began our work, you allowed me to steer the thesis towards a topic that I was interested in, even though it was outside of your comfort zone. Richard, thank you for spurring on much of the creative side of the thesis. You helped me bounce many ideas for possible applications of random graphs to economics, which really shaped the latter half of the work I did. I am very grateful for all of your input and am looking forward to working with you next year.

To Mum and Dad Thornton and Mum and Dad Leyland, thank you for championing me and taking the time to listen to all of my thoughts. Thank you especially for looking after my wife, my first daughter, and my second daughter who will be arriving soon. I know that your support has meant the world to Chelsea, and so it means the world to me. A special thanks goes out to you Dad for pioneering the

way by being the first in our family to complete your PhD thesis, and showing me that it was possible to embark on an ambitious research project. Albeit I am not quite up to your calibre, though I am hot on your tail.

Finally, to my best friend Connor. Thank you for all of the political, social, and psychological discussions which broke up the moments of monotony that inevitably ensue. I am grateful that you consistently remind me to ground my work in reality, and offer useful suggestions even though you are in an entirely different field of research. Moreover, thank you for putting me onto a podcast that would highlight to me the importance of diffusion, thereby confirming my choice of application for the thesis.

Abstract

The study of random graphs has become of central importance in both combinatorics, and mathematical modelling of real-world networks. Suppose that you have a coin which, when flipped, lands on heads with probability $p \in [0, 1]$, and tails with probability $1 - p$. Given a set of n vertices, construct a graph on them as follows. For each possible edge between the n vertices you toss the coin, including that edge in the graph if the coin lands heads. Such a construction is called the binomial random graph and its asymptotic properties (as $n \to \infty$) have been well studied when $p = p(n)$ is a sequence of probabilities depending on n. In this thesis, we present the proof of the classic result that $p = 1/n$ is a sharp threshold for the emergence of the giant component in the binomial random graph. We then turn to the question of whether such a threshold exists in graphs where the degree of each vertex is specified. To generate graphs with specified degrees, we use an algorithm known as the configuration model. Using the configuration model, a simple condition is presented for the existence of the giant component in simple graphs and multigraphs with an arbitrary degree distribution. The study of graphs with arbitrary degree distributions can be done in two ways, referred to as "canonical" and "microcanonical" ensembles. We discuss both the canonical and microcanonical ensembles, obtaining an exponential bound on the deviation of component sizes from the extinction probabilities of branching processes. This concentration result is derived within the canonical ensemble, and coalesces ideas from physics and mathematics. Finally, we look at a recently developed model of information cascades on networks. We analyse the model in light of our work, and suggest some avenues for future research.

Contents

Chapter 1 Introduction 1

Chapter 2 Background 4
- 2.1 Graphs . . . 4
- 2.2 Probability . . . 6
- 2.3 Asymptotics . . . 13
- 2.4 Random Graphs . . . 15
  - 2.4.1 Binomial Random Graphs . . . 15
  - 2.4.2 Erdős -Rényi Random Graphs . . . 16
  - 2.4.3 Random subsets . . . 16
  - 2.4.4 Two-round exposure (coupling) . . . 17
- 2.5 Graph Properties . . . 17
  - 2.5.1 Monotone Graph Properties . . . 18
  - 2.5.2 Thresholds . . . 19
- 2.6 Inequalities . . . 20

Chapter 3 The Giant Component in Erdős-Rényi Random Graphs 25
- 3.1 Branching Processes . . . 26
  - 3.1.1 Uniform (Galton-Watson) Branching Process . . . 26
  - 3.1.2 Generating Functions and Uniform Branching Processes . . . 27
  - 3.1.3 Extinction Probability . . . 28
  - 3.1.4 Exploration of Branching Processes . . . 31
  - 3.1.5 Branching Processes on $\mathcal{G}_{n,p}$ . . . 33
- 3.2 Threshold for existence of the Giant Component . . . 35

Chapter 4 The Giant Component for an Arbitrary Degree Distribution 43
- 4.1 Degree Distributions . . . 43

4.1.1 Degree Distribution for $\mathcal{G}_{n,p}$ 44
4.2 The Configuration Model 45
4.3 The Giant Component in The Configuration Model 49
4.3.1 Generating functions 51
4.3.2 Component sizes 53
4.3.3 The threshold for the emergence of the giant component . . 56
4.3.4 Formal analysis of BRP's with the configuration model . . . 59
4.3.5 Average number of t-th neighbours 66
4.3.6 Generating Functions for the Giant Compononent 67
4.3.7 Limitations and Extensions 69

Chapter 5 Information Cascades and Diffusion Games 72
5.1 Single-Type Diffusion Games 73
5.2 Analysis of Single-Type Diffusion Games 77
5.2.1 Mapping Strategies to Outcomes 77
5.2.2 Limit Beliefs and Viral Inference 82
5.2.3 Viral inference with knowledge of time 84

Chapter 6 Concluding Remarks 86

References 88

Chapter 1

Introduction

The theory of random graphs was introduced by Erdős & Rényi in 1959 [18, 19] and its study has become widespread among combinatorialists, physicists, and economists since. A key goal of Erdős & Rényi was to understand how the structure of a random graph changed as one varied the expected number of edges in the graph. In particular, they studied the size of connected *components* of the random graph. In this pursuit, Erdős & Rényi discovered a remarkable "double-jump" in the size of the largest component, the order of which increased from $O(\log n)$ to $O(n^{2/3})$, and then to $O(n)$, as one increased the expected number of edges in the graph. The component of size $O(n)$ was appropriately called *the giant component*. The intricacy of the graph defined by Erdős and Rényi became immediately apparent to the mathematical community, and as such, random graph theory is still a major area of interest.

There were two directions in which researchers proceeded from the results in [18,19]. The first direction was to analyse the subtle changes in component size associated with the double-jump, a question which was finally settled by Łuczak [38]. The second direction was to dream up new types of random graphs, and see whether one observed the same behaviour in the component sizes. A very general model of random graphs known as the *configuration model* was developed independently by Bollabás and Wormald [8, 57]. It was proved by Molloy & Reed [42, 43] that this model did in fact display the same behaviour as the simple model developed by Erdős and Rényi.

Ever since these major development in the theory of random graphs [8, 42, 43, 57], many improvements and alternative proofs have been offered to the results of [42,43].

Due to its remarkable overlap with a field of research in physics known as *percolation theory*, physicists have offered several contributions to the field of random graphs. We look at a key contribution of physicist M. Newman [47] to the results of [42, 43].

The authors of [18, 19] used both combinatorial and probabilistic arguments to prove their results. As the field of *probabilistic combinatorics* has emerged, the theory of random graphs has demanded a broader skill set than probability and combinatorics alone. Indeed, many techniques used in random graph theory today draw on deep results from measure theory, functional analysis and even complex analysis. Among the development of these new tools in random graph theory, older tools such as *probability generating functions* and *branching processes* [54] have re-emerged, demonstrating themselves to still be useful for proving results which are at the frontier of research. This thesis focuses primarily on these classical methods and their use in proving recent results.

The model developed in [8, 57] turned out to be extremely useful for describing real-world networks. A field of study emerged which looked at processes occurring on networks. This has been particularly important in recent economic research.

The modern study of economics is built on *models*: simplified mathematical descriptions of reality that capture important details of interest. With the growing empirical evidence that networks of relationships play a key function in economies (see for example [44]), economists have found themselves needing more advanced techniques to model the formation, effects, and efficiency of complex networks [30]. A major question in studying the effects of networks on economics is *what kind of processes occur in a network, and how can we model them precisely?* This is a question of *diffusion on networks*. Models of diffusion which do not incorporate network structure have been around since the 18th century.

One of the earliest models of diffusion was developed by Bernoulli after controversy broke out in France with regards to inoculation against smallpox [16]. Various models of disease transmission have been developed since Bernoulli, (see [20] for a history of such models), though many of the models developed did not take into account any explicit network structure.

A major breakthrough in the study of diffusion was the Susceptible-Infective (SI), Susceptible-Infective-Susceptible (SIS), and Susceptible-Infective-Removed (SIR) models, originally designed to describe the spread of infectious diseases [4]. These

models were able to explicitly capture aspects of the structure of the network on which the diffusion took place. These have become particularly important in recent times, with viral bacterial infections known as "superbugs" poised to become the biggest worldwide cause of death by 2050 [53]. Newman, the author of a heuristic argument [47] which we will look at in Chapter 4, has an excellent paper on SIR models of diffusion, which uses exactly the same methods which we apply in Chapter 4. Watts, a co-author of [47], used these same methods to develop a model for the spread of information on networks which addressed an important class of problems in economics known as *binary decisions with externalities* [51]. Watts called his work "a simple model of global cascades" [55], and the study of *information cascades* on networks was born. Models of information cascades have important implications for the success of marketing campaigns [36], the spread of rumours [6], the spread of new products and services [37], and even in improving welfare in developing countries [5].

In this thesis, we will introduce the fundamental ideas and results established in relation to the emergence of the giant component in random graphs. We will begin in with the necessary definitions and notation needed to understand the methods used in the thesis (Chapter 2), before providing an introduction to branching processes and proving the main results of Erdős and Rényi (Chapter 3). We then generalise the main result of Chapter 3 to graphs with an arbitrary degree sequence (Chapter 4). Finally, we look at an application of our research to models of information cascades, and suggest some avenues for future research (Chapter 5).

Chapter 2

Background

We now present several important concepts and definitions which will be pertinent to the results of this thesis. The main references for this chapter are Diestel [15, Chapter 1] for Section 2.1, Janson, Łuczak & Rucinski [32, Chapter 1] for Sections 2.3, 2.4, 2.4.1 to 2.4.4 and 2.5 and Frieze & Karoński [22, Chapters 21, 22] for Sections 2.2 and 2.6 respectively.

2.1 Graphs

The most fundamental construct throughout this thesis will be the idea of a *graph*. We make a few notes on notation before presenting some definitions. We denote the natural numbers, positive integers, real numbers, and complex numbers by $\mathbb{N}$, $\mathbb{Z}^+$, $\mathbb{R}$, and $\mathbb{C}$ respectively, and we write $[n]$ for the set of the first n positive integers $\{1, 2, \ldots, n\} \subseteq \mathbb{Z}^+$. Given a set X, and a positive integer $k \leq |X|$, we denote by $\binom{X}{k}$ the set of k-element subsets of X.

Definition 2.1. A *graph* is a pair $G = (V, E)$, where V is a finite set, and $E \subseteq \binom{V}{2}$.

To avoid ambiguity, we will always assume tacitly that $V \cap E = \emptyset$. We will say that a graph $G = (V, E)$ is a graph *on* V. The elements of V are called the *vertices* of the graph, and elements of E the *edges*. Generally we will deal with the case where $V = [n]$. We usually write ij for an edge $\{i, j\}$ between vertices i and j in V. Given a graph $G = (V, E)$, we say that two distinct edges $e_i, e_j \in E$ are *incident* if there exists $v \in V$ such that $v \in e_i$ and $v \in e_j$, that is, e_i and e_j share a common vertex. We say that two distinct vertices $v_i, v_j \in V$ are *adjacent* if $v_i v_j \in E$, that is, if there is an edge between v_i and v_j. If v_i and v_j are adjacent we call v_j a *neighbour* of v_i.

Definition 2.2. The *degree* of a vertex $v \in V$, denoted $\deg_G(v)$, is the number of edges which are incident with v in G. That is,

$$\deg_G(v) = |\{e \in E \colon e \text{ is incident with } v\}|.$$

When it is clear which graph we are referring to, we simply write $\deg(v)$ rather than $\deg_G(v)$. In a *simple* graph, as defined in Definition 2.1, a vertex can have degree at most $n-1$. Instead, one could define a *multigraph*.

Definition 2.3. A *multigraph* is a pair (V, E) of disjoint sets (of *vertices* and *edges*) together with a map $E \to V \cup \binom{V}{2}$, assigning to every edge either one or two vertices. If $e \mapsto v$, then we say that e is a *loop* at v, which is an edge from v to itself.

In both *graphs* and *multigraphs*, two vertices $i, j \in V$ have an edge between them if and only if there is an edge between $j, i \in V$. One obtains a *directed* graph by placing an *orientation* on the edges of a graph $G = (V, E)$, such that an edge ij is distinguished from edge ji. We also note that the vertices of a graph are distinguishable from each other since they come from a set. This property defines a *labelled* graph.

Remark 2.4. All graphs throughout this thesis will be assumed to be *simple, undirected, and labelled* in accordance with Definition 2.1.

We will need to know a few other basic concepts in order to understand the analysis of random graphs later on. We present some fundamental definitions below.

Definition 2.5. Let $G = (V, E)$ be a graph. A *subgraph* of G is a graph $H = (W, F)$ such that $W \subseteq V$ and $F \subseteq E$.

Definition 2.6. A *path* on G is a sequence of distinct vertices $(v_0, v_1, \ldots, v_k)$ with $k \in \mathbb{N}$, such that $v_i v_{i+1} \in E$ for $i = 1, 2, \ldots, k-1$. The *length* of this path is k (the number of edges).

One can think of a path in the natural way as following a series of edges from one vertex to another, that is, following incident edges to adjacent vertices. Adding an edge joining the endvertices of a path of length ≥ 3 gives a *cycle*.

Definition 2.7. A *cycle* on G is a sequence of vertices $(v_1, \ldots, v_k, v_1)$ such that $(v_1, \ldots, v_k)$ is a path, and $v_k v_1 \in E$. The *length* of the cycle is k.

It is useful to define a notion of distance on a graph. This is done in the natural way.

Definition 2.8. Let v, w be distinct vertices in a graph G. The (graph) *distance* between v and w, denoted $d_G(v, w)$, is the length of the shortest path between them. If there is no path between v and w then we set $d_G(v, w) = \infty$.

We use the above two definitions to introduce the notion of connectedness of a graph. We also define *components* of a graph, which will be essential to the statement of our main theorems.

Definition 2.9. A graph is $G = (V, E)$ is *connected* if there is a path from v to w in G, for all $v, w \in V$. Equivalently, G is connected if $d_G(v, w) < \infty$ for all $v, w \in V$. A maximally connected subgraph of G is called a *component* (or *connected component*) of G. Components are nonempty.

A connected graph without any cycles is called a *tree*. Trees will be an important object of study when we locally approximate the structure of a random graph (see Theorems 3.10 and 3.11 and Lemma 4.12).

Next, we introduce some basic probability theory and asymptotic notions before we introduce the main objects of study in this thesis, random graphs.

2.2 Probability

We use $\mathrm{Bin}(n, p)$, $\mathrm{Be}(p)$, $\mathrm{Po}(\lambda)$, and $\mathrm{N}(\mu, \sigma^2)$ to denote the Binomial, Bernoulli, Poisson, and Normal distributions respectively. We write $X \sim \mathcal{D}$ to mean that X is a random variable with distribution $\mathcal{D}$, (for example, $X \sim \mathrm{N}(0, 1)$). We begin by introducing the standard notions of conditional probability, discrete random variables, conditional expectation, and independence. We will write $\mathbb{P}$ for probability.

Definition 2.10. Let A and B be two events defined on the same probability space Ω. The probability of A given B, is defined by

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

We will only need to work with discrete probability spaces in this thesis. Hence when we write Ω for a probability space, we really mean the triple $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F} = 2^{\Omega}$ is the set of all subsets of Ω and $\mathbb{P}\colon \mathcal{F} \to [0,1]$ is a probability measure.

Definition 2.11. A *discrete random variable* is a function $X\colon \Omega \to E$, from a countable sample space Ω to a countable set E. A discrete random variable can be described entirely in terms of its *probability mass function* (p.m.f.) $f_X\colon E \to [0,1]$, which is defined by

$$f_X(k) = \mathbb{P}(X = k) = \mathbb{P}(\{\omega \in \Omega\colon X(\omega) = k\}).$$

Definition 2.12. Let X be a discrete random variable, and let H be an event, with both X and H defined on the same probability space Ω. The expectation of X given H, is defined by

$$\mathbb{E}(X \mid H) = \sum_{k \in H} k\mathbb{P}(X = k \mid H).$$

Definition 2.13. Two events variables A and B defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ are said to be *independent* if $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.

An important case of the above definition is when one has two random variables X and Y on Ω. Then the events "$X = x$" and "$Y = y$" are independent if

$$\mathbb{P}(X = x \mid Y = y) = \mathbb{P}(X = x).$$

A sequence of random variables $\{X_n\}_{n \in \mathbb{N}}$ is *independent and identically distributed* (i.i.d.) if X_n has the same probability distribution for all $n \in \mathbb{N}$, and X_n is independent of X_m for all $n \neq m$.

We denote by $\mathbb{1}_{[\mathcal{E}]}$ the indicator function of the event $\mathcal{E}$, which equals 1 if event $\mathcal{E}$ occurs and 0 otherwise. We will often consider random variables which are indicator functions of some event. Such variables clearly have Bernoulli distribution with $p = \mathbb{P}(\mathcal{E})$. The expected value and variance of a random variable X (if they exist) will be denoted by $\mathbb{E}(X)$ and $\mathrm{Var}(X)$ respectively. We have the following extremely important results.

Lemma 2.14 (Markov's Inequality)**.** *Let X be a non-negative random variable. Then, for all $t > 0$,*

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

Proof. We have that

$$X = X\,\mathbb{1}_{[X \geq t]} + X\,\mathbb{1}_{[X < t]} \geq X\,\mathbb{1}_{[X \geq t]} \geq t\,\mathbb{1}_{[X \geq t]}.$$

Hence

$$\mathbb{E}(X) \geq \mathbb{E}(t\mathbb{1}_{[X \geq t]}) = t\,\mathbb{E}(\mathbb{1}_{[X \geq t]}) = t\mathbb{P}(X \geq t).$$

□

Corollary 2.15 (First Moment Method)**.** *Let X be a non-negative integer-valued random variable. Then,*

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

Proof. Put $t = 1$ in Markov's Inequality. □

Corollary 2.16 (Chebyshev Inequality)**.** *If X is a random variable with a finite mean and variance, then for $t > 0$,*

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

Proof. By Markov's Inequality,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) = \mathbb{P}((X - \mathbb{E}(X))^2 \geq t^2) \leq \frac{\mathbb{E}\,(X - \mathbb{E}(X))^2}{t^2} = \frac{\mathrm{Var}(X)}{t^2}.$$

□

We will also use the law of total expectation, which states that for any finite or countable partition of the probability space $\Omega = \mathcal{E}_1 \cup \mathcal{E}_2 \cup \ldots$ and any random variable X defined on Ω,

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} \mathbb{E}(X \mid \mathcal{E}_i)\,\mathbb{P}(\mathcal{E}_i).$$

In particular, if $X = \mathbb{1}_{[\mathcal{E}]}$ then $\mathbb{P}(\mathcal{E}) = \sum_{i=1}^{\infty} \mathbb{P}(\mathcal{E} \mid \mathcal{E}_i)\,\mathbb{P}(\mathcal{E}_i)$.

Lemma 2.17 (Union bound)**.** *For any finite or countable set of events $\mathcal{E}_1, \mathcal{E}_2, \ldots$ of the probability space Ω, we have*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} \mathcal{E}_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(\mathcal{E}_i).$$

A useful tool which we make repeated use of in throughout the thesis is *probability generating functions* (see [26, § 5.1]). Probability generating functions form the basis for analysing the properties of *branching processes*, which we will introduce in Section 3.1.

Definition 2.18 (Probability Generating Function)**.** Let X be a discrete random variable taking on values in the non-negative integers $\{0, 1, 2, \ldots\}$. The *probability generating function* (p.g.f.) $f_X \colon \mathbb{R} \to \mathbb{R}$ of X is defined as

$$f_X(z) = \mathbb{E}(z^X) = \sum_{k=0}^{\infty} \mathbb{P}(X = k) z^k, \quad \text{for all } z \in \mathbb{R}.$$

Remark 2.19. *Because the coefficients of z^k are all between 0 and 1, and sum to 1, the above power series converges absolutely for all $z \in \mathbb{R}$ with $|z| \leq 1$, though the radius of convergence can be larger. By definition, the p.g.f. is uniquely determined by the distribution of X. The p.g.f. also uniquely determines the distribution of X. This is because we have that $\mathbb{P}(X = k) = [z^k] f_X(z)$, where $[z^k] f_X(z)$ denotes the coefficient of z^k in $f_X(z)$.*

We state here a few simple properties of the probability generating function.

Lemma 2.20. *Let X be a discrete random variable taking on values in the non-negative integers, and let f_X its denote its probability generating function as above. Then*

(i) $f_X(0) = \mathbb{P}(X = 0)$,

(ii) $f_X(1) = 1$,

(iii) f_X *is continuous on* $[0,1]$,

(iv) f_X *is non-decreasing on* $[0,1]$,

(v) f_X *is convex on* $[0,1]$.

Proof.

(i) $f_X(0) = \sum_{k=0}^{\infty} \mathbb{P}(X=k)\, 0^k = \mathbb{P}(X=0)$.

(ii) $f_X(1) = \sum_{k=0}^{\infty} \mathbb{P}(X=k)\, 1^k = \sum_{k=0}^{\infty} \mathbb{P}(X=k) = 1$.

(iii) Recall that a power series is continuous inside its radius of convergence. We noted in Remark 2.19 that the radius of convergence for a probability generating function is at least 1. Hence f is continuous on $[0,1]$.

(iv) We have that

$$f'(z) = \sum_{x=1}^{\infty} \mathbb{P}(X=k) k z^{k-1} \geq 0$$

for any $z \in [0,1]$.

(v) We have that

$$f''(z) = \sum_{k=2}^{\infty} \mathbb{P}(X=k) k(k-1) z^{k-2} \geq 0$$

for any $z \in [0,1]$.

□

Consider the following example of a p.g.f. which will be useful for our later analysis.

Example 2.21. *Let* $X \sim \text{Bin}(n,p)$. *Then for any* $k = 0, 1, \ldots, n$, *we have that* $\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$. *Hence*

$$\begin{aligned} f_X(z) &= \sum_{k=0}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} z^k \\ &= \sum_{k=0}^{\infty} \binom{n}{k} (pz)^k (1-p)^{n-k} \\ &= (pz + 1 - p)^n. \qquad \text{(by the Binomial Theorem)} \end{aligned}$$

Hence $f_X(z) = (pz + 1 - p)^n$ *for all* $z \in \mathbb{R}$.

A natural question is how the p.g.f. behaves under sums and products of random variables. We make note here of the following useful theorem.

Theorem 2.22. *Let* $X_1, \ldots, X_n$ *be a sequence of independent and identically distributed random variables with a common p.g.f.* f_X*. Let* Y *be a random variable independent of* $X_1, \ldots, X_n$ *with p.g.f.* f_Y*, and let* $Z_Y = X_1 + \cdots + X_Y = \sum_{k=1}^{Y} X_k$*. Then the p.g.f of* Z_Y *is*

$$f_{Z_Y}(z) = f_Y\left(f_X(z)\right).$$

Proof. By direct calculation,

$$\begin{aligned}
f_{Z_Y}(z) = \mathbb{E}(z^{Z_Y}) &= \sum_{k=0}^{\infty} \mathbb{E}\left(z^{Z_Y} \mid Y = k\right) \mathbb{P}(Y = k) && \text{(conditioning on Y)} \\
&= \sum_{k=0}^{\infty} \mathbb{E}\left(z^{X_1 + \cdots + X_k}\right) \mathbb{P}(Y = y) \\
&= \sum_{k=0}^{\infty} \mathbb{E}\left(z^{X_1}\right) \cdots \mathbb{E}\left(z^{X_k}\right) \mathbb{P}(Y = k) && \text{(independence)} \\
&= \sum_{k=0}^{\infty} \left(f_X(z)\right)^k \mathbb{P}(Y = k) \\
&= f_Y\left(f_X(z)\right),
\end{aligned}$$

as required. □

Corollary 2.23. *Let* $X_1, \ldots, X_n$ *be as above and let* Y *be a random variable with probability distribution* $\mathbb{P}(Y = m) = 1$ *for some positive integer* m*. Then the probability generating function of* $Z = X_1 + \cdots + X_m$ *is* $f_Z(z) = f_X(z)^m$.

Definition 2.24 (Stochastic Dominance)**.** Let X, Y be random variables with outcomes in the same measurable space E. We say that X *stochastically dominates* Y if

$$\mathbb{P}(X \geq x) \geq \mathbb{P}(Y \geq x), \quad \text{for all } x \in E,$$

and for at least one $x \in D$,

$$\mathbb{P}(X \geq x) > \mathbb{P}(Y \geq x).$$

Stochastic dominance provides a partial order on probability distributions with outcomes in E. We write $Y \preceq X$ if X stochastically dominates Y.

Remark 2.25. *The idea of stochastic dominance is also important in economics in the area of decision theory. The type of stochastic dominance described above is referred to as first-order stochastic dominance, and is equivalent to stating that for two "gambles" with distributions X and Y over some set of outcomes E, every expected utility maximiser with an increasing utility function prefers gamble X over gamble Y. There are concepts of higher order stochastic dominance in this literature as well [27].*

Example 2.26. *Let $m, n \in \mathbb{N}$ such that $m < n$ and let $0 < p < 1$ be a real number. If $Y \sim \text{Bin}(m, p)$ and $X \sim \text{Bin}(n, p)$ then $Y \preceq X$; that is, X stochastically dominates Y. This is clear since for example, $p^n = \mathbb{P}(X \geq n) > \mathbb{P}(Y \geq n) = 0$ In particular, if $Y \sim \text{Bin}(n-k, p)$ for some integer $0 < k < n$, then $Y \preceq X$. This will be useful later on when we analyse branching processes.*

We also state without proof the following fact about sums of independent binomial random variables.

Lemma 2.27. *Let $0 < p < 1$ and let $X_1, \ldots, X_n$ be a sequence of random variables with $X_j \sim \text{Bin}(n_j, p)$ for positive integers n_j, where $j = 1, 2, \ldots, N$. Then*

$$\sum_{j=1}^{N} X_j \sim \text{Bin}\Big(\sum_{j=1}^{N} n_j, p\Big). \tag{2.1}$$

We now introduce a modern technique in probability theory known as *coupling*. This method is essential for one important proof in Chapter 4, and is also used in a minor result in Section 2.5.1. The following definition can be found in [29].

Definition 2.28. Let X and Y be random variables on the same sample space Ω. A *coupling* of X and Y is any ordered pair of random variables (X', Y') taking

values in $\Omega \times \Omega$, whose marginals have the same distribution as X and Y. That is,

$$X' \stackrel{D}{=} X, \qquad Y' \stackrel{D}{=} Y,$$

with $\stackrel{D}{=}$ denoting equality in distribution.

We will return to this definition shortly, once we have introduced some asymptotics.

2.3 Asymptotics

We use the following standard notation for the asymptotic behaviour of the relative order of magnitude of two sequences of numbers a_n and b_n, depending on a parameter $n \to \infty$. All asymptotics in this thesis are as $n \to \infty$ unless otherwise specified.

- $a_n = O(b_n)$ as $n \to \infty$ if there exists constants C and n_0 such that $|a_n| \leq C|b_n|$ for $n \geq n_0$. That is, if the sequence $|a_n|/|b_n|$ is bounded, except possibly for some small values of n for which the ratio may be undefined.

- $a_n = \Omega(b_n)$ as $n \to \infty$ if there exists constants $c > 0$ and n_0 such that $|a_n| \geq c|b_n|$ for $n \geq n_0$. This is equivalent to $b_n = O(a_n)$.

- $a_n = \Theta(b_n)$ as $n \to \infty$ if $a_n = O(b_n)$ and $a_n = \Omega(b_n)$. That is, there exists constants $C, c > 0$ and n_0 such that $c|b_n| \leq |a_n| \leq C|b_n|$ for $n \geq n_0$. This is sometimes expressed by saying that a_n and b_n are of the *same order of magnitude*.

- $a_n = o(b_n)$ as $n \to \infty$ if $a_n/b_n \to 0$, that is, if for every $\epsilon > 0$, there exists n_ϵ such that $|a_n| < \epsilon b_n$ for $n \geq n_\epsilon$.

We also introduce here notions of probability asymptotics. Let $\{\Omega_n\}_n$ be a sequence of probability spaces with parameter n. An event $\mathcal{E}_n \subseteq \Omega_n$ holds *with high probability* (abbreviated w.h.p.), if $\mathbb{P}(\mathcal{E}_n) \to 1$ as $n \to \infty$. We follow the convention of Janson, Łuczak & Rucinski [32, § 1.2] and introduce a probabilistic version of the little-oh notation. Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables and $\{a_n\}_{n\in\mathbb{N}}$ a sequence of positive real numbers. We then define

- $X_n = o_p(a_n)$ as $n \to \infty$ if for every $\epsilon > 0$, *w.h.p.* $|X_n| < \epsilon a_n$.

We will use this particular notation in our statement of Theorem 3.11.

There will be an interplay between both probability and asymptotics throughout the thesis. We now define what it means for a sequence of random variables to converge in *distribution* and in *probability*, after which we return to the idea of coupling from Section 2.2 (see Definition 2.28). The following two definitions are from [32, § 1.2].

Definition 2.29. A sequence $\{X_n\}_{n\in\mathbb{N}}$ of real-valued random variables is said to *converge in probability* to a random variable X (denoted $X_n \xrightarrow{p} X$)) if for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X - X_n| > \epsilon) = 0.$$

Definition 2.30. A sequence $\{X_n\}_{n\in\mathbb{N}}$ of real-valued random variables is said to *converge in distribution* to a random variable X (denoted $X_n \xrightarrow{d} X$)) if

$$\lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$$

for every real x at which $\mathbb{P}(X \leq x)$ is continuous.

In some circumstances it is useful to know when two sequences of random variables are essentially "the same" as $n \to \infty$.

Definition 2.31. Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of non-negative integer-valued random variables converging in probability to X, and $\{Y_n\}_{n\in\mathbb{N}}$ a sequence of non-negative integer-valued random variables converging in probability to Y. A sequence of couplings (X_n, Y_n) is *good* if

$$\sum_{k=0}^{\infty} \mathbb{P}(X'_n = k, Y'_n = k) = 1 - o(1),$$

where X' and Y' denote the marginal distributions of X and Y respectively.

The important use of the above definition is going to be when $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ are random variables on $\Omega_n = 2^{\binom{n}{2}}$, the set of all graphs on n vertices, in which case we are constructing an isomorphism of random graphs. Note that we write $2^{\binom{n}{2}}$ rather than $2^{\binom{[n]}{2}}$ for the set of all graphs on n vertices.

2.4 Random Graphs

We begin our discussion of random graphs by introducing the model established by Paul Erdős in 1947 [17]. At the time, Erdős did not explicitly call the structure that he had defined a "random graph". It was not for another twelve years that he, and Hungarian mathematician Alfréd Rényi formally introduced random graphs, and provided a remarkable analysis of their structure [18, 19]. The questions which Erdős & Rényi answered are the same questions we will look at in Chapter 3.

Let $[n] = \{1, 2, \ldots, n\}$ be a set of n vertices. We denote the set of all $2^{\binom{n}{2}}$ graphs on $[n]$ by Ω_n, and the family of all subsets of Ω_n by $\mathcal{F}_n$. Then one can describe the model introduced by Erdős as the probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P})$, where for every $\omega \in \Omega_n$,

$$\mathbb{P}(\omega) = 2^{-\binom{n}{2}}.$$

Another way to think of this is as $\binom{n}{2}$ independent coin tosses of a fair coin, one for each potential edge ij, where the outcome "heads" corresponds to including the edge ij in the graph, and the outcome "tails" corresponds to excluding the edge ij from the graph. Generally speaking, we are interested in the probability distribution induced by $\mathbb{P}$ on the family of graphs $\mathcal{F}_n$, and we rarely differentiate between two graphs with the same such distribution. However, it is not sufficient to formally define a random graph as a probability distribution only; we will see there are some examples such as the "two-round exposure" technique (Section 2.4.4) which consider several random graphs at the same time.

2.4.1 Binomial Random Graphs

There are two main models of random graphs whose study has been most exhaustive, introduced in the next two subsections. Let p be a real number such that $0 \leq p \leq 1$. Then for a graph $G \in \Omega_n$ on a vertex set $[n]$ with $|E(G)| = e_G \leq \binom{n}{2}$ edges, let the probability of G be

$$\mathbb{P}(G) = p^{e_G}(1-p)^{\binom{n}{2}-e_G}.$$

This is called the *binomial random graph*, denoted by $\mathcal{G}_{n,p}$. Equivalently one can think of this as the result of a process whereby we begin with an empty graph on $[n]$ vertices and undertake $\binom{n}{2}$ independent coin tosses, one for each potential edge ij where the probability of "heads" is p, corresponding to the probability of adding edge ij to the graph. For $p = \frac{1}{2}$ we have exactly the model introduced by Erdős in 1947, though $\mathcal{G}_{n,p}$ was introduced by Gilbert [24] in 1959, contemporaneously

and independently of the work of Erdős & Rényi [18]. The results of Chapter 3 are primarily based upon the binomial random graph.

2.4.2 Erdős -Rényi Random Graphs

In many cases we may be more interested in a situation where the number of edges of a binomial random graph $\mathcal{G}_{n,p}$ is fixed. If we condition the binomial random graph on the event that $|E(\mathcal{G}_{n,p})| = M$, we arrive at the *Erdős -Rényi random graph*, denoted by $\mathcal{G}_{n,M}$. This was the model introduced by Erdős and Rényi in 1959 [18] which we mentioned at the beginning of Section 2.4. The graph $\mathcal{G}_{n,M}$ is sometimes called the *uniform random graph*. Given an integer M such that $0 \leq M \leq \binom{n}{2}$, define Ω to be the family of all graphs on $[n]$ with exactly M edges, and for any graph $G \in \Omega$, let

$$\mathbb{P}(G) = \binom{\binom{n}{2}}{M}^{-1}.$$

We will usually capitalize the second subscript when referring to an Erdős -Rényi random graph, to avoid confusion with the binomial random graph $\mathcal{G}_{n,p}$. One may think of this as the process in which we begin with an empty graph on $[n]$ vertices and insert M edges to the graph in such a way that all possible $\binom{\binom{n}{2}}{M}$ choices are equally likely. In fact, when $\binom{n}{2}p$ is (loosely speaking) close to M, the binomial and Erdős -Rényi models are in a very precise sense 'equivalent' for large n (see [32, Proposition 1.12, 1.13].

2.4.3 Random subsets

It is often convenient to analyse random graphs in the more general context of random subsets of a set. The proofs of properties of monotonicity, equivalence, and thresholds (which we will introduce shortly) are often nearly identical, and we achieve a higher level of generality for no additional cost. The $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,M}$ random graphs introduced already fall quite nicely into this framework.

Let Γ be a finite set with $|\Gamma| = N$, and let p be a real number with $0 \leq p \leq 1$. Then we obtain the random subset Γ_p of Γ as follows. For each element of Γ, independently flip a coin which lands heads with probability p, and add only those elements to Γ_p for which the coin lands heads. The elements for which the coin flip lands tails are not included in Γ_p. The distribution of Γ_p is the probability distribution on $\Omega = 2^\Gamma$ given by $\mathbb{P}(F) = p^{|F|}(1-p)^{|\Gamma|-|F|}$ for $F \subseteq \Gamma$. Similarly if we let M be an integer with $0 \leq M \leq N$, then we obtain Γ_M by randomly choosing an element of $\binom{\Gamma}{M}$; so Γ_M has uniform distribution $\mathbb{P}(F) = \binom{N}{M}^{-1}$ for $F \in \binom{\Gamma}{M}$.

Example 2.32. *Let* $\Gamma = \binom{[n]}{2}$*, then one can think of* Γ *as all possible edge pairs on the vertex set* $[n]$*. Hence in this case we obtain* $\mathcal{G}_{n,p} = \Gamma_p$ *and* $\mathcal{G}_{n,M} = \Gamma_M$*.*

2.4.4 Two-round exposure (coupling)

The two-round exposure is a useful technique which constructs the binomial random graph in two independent stages. Suppose that $p_1 < p_2$, and define p_0 by

$$1 - p_2 = (1 - p_1)(1 - p_0),$$

or equivalently

$$p_0 = \frac{p_2 - p_1}{1 - p_1}.$$

Then an element of Γ is absent from in the random subset Γ_{p_2} if it is absent from both Γ_{p_0} and Γ_{p_1}. It follows that

$$\Gamma_{p_2} = \Gamma_{p_0} \cup \Gamma_{p_1},$$

where the two random subsets Γ_{p_0} and Γ_{p_1} are independent. In the case of random graphs, we first generate $\mathcal{G}_{n,p_0}$, and then independently generate $\mathcal{G}_{n,p_1}$ on the same vertex set. We obtain $\mathcal{G}_{n,p_2}$ by taking the union of the edge sets. Intuitively, the first round "exposes" some edges, and the second round "exposes" some more edges. The two-round exposure technique is really a kind of isomorphic coupling (see Definitions 2.28 and 2.31) of the random variable Γ_{p_2} with the jointly distributed random variables Γ_{p_0} and Γ_{p_1}.

2.5 Graph Properties

Although we have defined our random graphs to be labelled (that is, on a labelled set of vertices), we are mainly interested in properties that are independent of such labellings, that is, those properties which are preserved by isomorphism. The following definition is from Diestel [15].

Definition 2.33. Let $G = (V, E)$, and $G' = (V', E')$ be two graphs. We say that G and G' are *isomorphic* (write $G \cong G'$) if there exists a bijective map $\phi\colon V \to V'$ such that $xy \in E(G)$ if and only if $\phi(x)\phi(y) \in E(G')$. We say ϕ is an *isomorphism*. If $G = G'$, then we say that ϕ is an *automorphism*.

A *graph property* (sometimes called a *graph invariant*) is a property which is preserved by isomorphism. We will denote a graph property by $\mathscr{P}$, and we identify

$\mathscr{P}$ with the corresponding family of all labelled graphs on the vertex set $[n]$ which have that property, that is, $\mathscr{P} \subseteq 2^{\binom{n}{2}}$.

2.5.1 Monotone Graph Properties

A family of subsets $\mathcal{Q} \subseteq 2^{\Gamma}$ is called *increasing* if $A \subseteq B$ and $A \in \mathcal{Q}$ imply that $B \in \mathcal{Q}$. A family of subsets is *decreasing* if its complement in 2^{Γ} is increasing, or, equivalently, if the family of the complements in Γ is increasing. A family which is either increasing or decreasing is called *monotone*. A family $\mathcal{Q}$ is *convex* if $A \subseteq B \subseteq C$ and $A, C \in \mathcal{Q}$ imply $B \in \mathcal{Q}$. In the case where $\Gamma = \binom{[n]}{2}$, we have that $2^{\Gamma} = \Omega_n$, and therefore any family $\mathcal{Q} \subseteq 2^{\Gamma}$ is a family of graphs. If a family of graphs $\mathcal{Q}$ is invariant under isomorphism, then it can be identified with a graph property.

We now present our first proof which follows from the two-round exposure technique introduced in Section 2.4.4.

Lemma 2.34 (Bollobás, 1979)**.** *Let $\mathcal{Q}$ be an increasing property of subsets of Γ, $0 \leq p_1 \leq p_2 \leq 1$, and $0 \leq M_1, M_2 \leq N$. Then*

$$\mathbb{P}(\Gamma_{p_1} \in \mathcal{Q}) \leq \mathbb{P}(\Gamma_{p_2} \in \mathcal{Q})$$

and

$$\mathbb{P}(\Gamma_{M_1} \in \mathcal{Q}) \leq \mathbb{P}(\Gamma_{M_2} \in \mathcal{Q}).$$

Proof. We begin with the first inequality. Let p_0 be defined by $1-p_2 = (1-p_1)(1-p_0)$, so $p_0 = (p_2-p_1)/(1-p_1)$. Then by the two-round exposure technique, we have that $\Gamma_{p_2} = \Gamma_{p_0} \cup \Gamma_{p_1}$. Therefore since $\Gamma_{p_1} \subseteq \Gamma_{p_0} \cup \Gamma_{p_1} = \Gamma_{p_2}$, and $\mathcal{Q}$ is increasing, we have that $\Gamma_{p_1} \in \mathcal{Q}$ implies that $\Gamma_{p_2} \in \mathcal{Q}$. Hence

$$\mathbb{P}(\Gamma_{p_1} \in \mathcal{Q}) \leq \mathbb{P}(\Gamma_{p_2} \in \mathcal{Q})$$

as required. For the second inequality, we can proceed via two-round exposure, or alternatively we can construct a sequence of random subsets $\{\Gamma_M\}_{M=1}^{N}$ by selecting the elements of Γ one by one in a random order. If we take Γ_M to be the M-th subset in this sequence, then $\Gamma_{M_1} \subseteq \Gamma_{M_2}$, and so as above, $\Gamma_{M_1} \in \mathcal{Q}$ implies that $\Gamma_{M_2} \in \mathcal{Q}$. The result follows. □

2.5.2 Thresholds

In our study of random graphs, we are usually interested in the limiting probability that a random graph G has a certain monotonic property $\mathscr{P}$. That is, we are interested in $\mathbb{P}(G \in \mathscr{P})$ as $n \to \infty$. For many graph properties, the limiting probability jumps suddenly from 0 to 1 as one increases the expected number of edges of the graph past some "threshold" value. This phenomenon was noticed by Erdős and Rényi in their early papers on random graphs [18, 19]. Probably the most striking fact about monotone properties is that they always exhibit a threshold. This was proved by Bollobás and Thomason [11] for arbitrary random subsets, and thus is also true for monotone properties of $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,M}$. The higher level of generality provided by the framework of random subsets of a set allows us to talk about thresholds in random graphs other than just $\mathcal{G}_{n,p}$ or $\mathcal{G}_{n,M}$, as we do in Section 4.3. We state the theorem here so that we can refer back to it.

Theorem 2.35 (Bollobás & Thomason, 1987)**.** *Every nontrivial monotone graph property exhibits a threshold.*

We now formally define a threshold for an increasing family of sets $\mathcal{Q}$, though one can easily restate Definitions 2.36 and 2.37 in terms of decreasing properties of sets. We follow the definitions of Frieze & Karoński [22], though we provide them in the setting of random subsets of a set, whereas the authors provide them specifically in relation to $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,M}$.

Definition 2.36. A function $p^* = p^*(n)$ is a *threshold* for a monotone increasing property $\mathcal{Q}$, if

$$\lim_{n\to\infty} \mathbb{P}(\Gamma_p \in \mathcal{Q}) = \begin{cases} 0, & \text{if } p/p^* \to 0 \\ 1, & \text{if } p/p^* \to \infty. \end{cases}$$

One can make an analogous definition for random subsets Γ_M.

Definition 2.37. A function $M^* = M^*(n)$ is a *threshold* for a monotone increasing property $\mathcal{Q}$, if

$$\lim_{n\to\infty} \mathbb{P}(\Gamma_M \in \mathcal{Q}) = \begin{cases} 0, & \text{if } M/M^* \to 0 \\ 1, & \text{if } M/M^* \to \infty. \end{cases}$$

Threshold functions are only unique up to multiplication by a positive constant. That is, if $\mathcal{Q}$ is a monotone increasing graph property, and $p^* = p^*(n)$ is a threshold for $\mathcal{Q}$, then so is Cp^* for any constant $C > 0$. Some monotone graph properties exhibit thresholds which are more sensitive to changes in the threshold function. Such thresholds are called *sharp thresholds*, which we will only define for random subsets Γ_p.

Definition 2.38. A function $p^* = p^*(n)$ is a *sharp threshold* for a monotone increasing property $\mathcal{Q}$, if for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(\Gamma_p \in \mathcal{Q}) = \begin{cases} 0, & \text{if } p/p^* \leq 1 - \epsilon \\ 1, & \text{if } p/p^* \geq 1 + \epsilon. \end{cases}$$

If $p^* = p^*(n)$ is a threshold for a monotone increasing property, we say that Γ_p exhibits a *phase transition* around p^*. This terminology comes from the literature in Chemistry and Statistical Physics, where for example one might study the phase transition of water from a gas to a liquid and to a solid.

2.6 Inequalities

We present here some useful inequalities that will be used throughout the paper.

$$1 + x \leq e^x \qquad \text{for all } x \in \mathbb{R}. \tag{2.2}$$

$$1 - x \geq e^{-x/(1-x)} \qquad 0 \leq x < 1. \tag{2.3}$$

We also need Stirling's approximation, which can also be restated in terms of upper and lower bounds on $n!$. We only require its use asymptotically.

$$n! = (1 + o(1))\sqrt{2\pi n}\left(\frac{n}{e}\right)^n, \qquad \text{for all } n \in \mathbb{N}. \tag{2.4}$$

The following asymptotic formulas will also be used at various points in the thesis.

$$\binom{n}{k} = (1 + o(1)) \frac{n^k}{k!}, \qquad \text{if } k^2 = o(n). \tag{2.5}$$

$$\frac{1}{1-\alpha} = 1 + O(\alpha), \qquad \text{if } \alpha = o(1). \tag{2.6}$$

Equation (2.6) is an immediate consequence of the fact that for all $\alpha \in \mathbb{R}$, with $|\alpha| < 1$, one has that $1/(1-\alpha) = 1 + \alpha + \alpha^2 + \cdots = 1 + O(\alpha)$.

We note here that all logarithms in this thesis will be assumed to be natural unless otherwise specified. Chernoff's Bounds will be needed to give exponentially decreasing tail bounds on the sums of independent binomially distributed random variables. The proof is omitted but can be found in Section 2.1 of Janson, Łuczak, & Rucinski [32].

Theorem 2.39 (Chernoff's Bounds)**.** *If $X \sim \mathrm{Bin}(n, p)$, and $\lambda = np$, then, with $\varphi(x) = (1+x)\log(1+x) - x$, $x \geq -1$ (and $\varphi = \infty$ for $x < -1$)*

$$\mathbb{P}(X \geq \mathbb{E}(X) + t) \leq \exp\left(-\lambda\varphi\left(\frac{t}{\lambda}\right)\right) \leq \exp\left(-\frac{t^2}{2(\lambda + t/3)}\right), \qquad t \geq 0; \tag{2.7}$$

$$\mathbb{P}(X \leq \mathbb{E}(X) - t) \leq \exp\left(-\lambda\varphi\left(\frac{-t}{\lambda}\right)\right) \leq \exp\left(-\frac{t^2}{2\lambda}\right), \qquad t \geq 0. \tag{2.8}$$

Hoeffding's Inequality provides us with an exponentially decreasing tail bound for sums of independent random variables in a more general setting. We will need this when providing bounds for sums of independent random variables which are not binomially distributed.

Theorem 2.40 (Hoeffding's Inequality)**.** *Let $X_1, \ldots X_n$ be independent random variables such that X_i is strictly bounded by $[a_i, b_i]$. Denote by $\overline{X}$ the average of the X_i's, that is, $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Then for all $t \geq 0$,*

$$\mathbb{P}(\left|\overline{X} - \mathbb{E}(\overline{X})\right| \geq t) \leq 2\exp\left(-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right). \tag{2.9}$$

In the thesis, I needed to apply Chernoff's bound Theorem 2.39 to a convergent sequence of binomial random variables $\{X_k\}_{k\in\mathbb{N}}$ and look at its limiting behaviour

(see Theorem 3.11). I noticed that it would be useful to be able to apply Theorem 2.39 to the asymptotic mean of the sequence. I therefore state and prove sufficient conditions under which this can be done.

Theorem 2.41 (Asymptotic Chernoff's Bounds (Original))**.** *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables with $X_n \sim \text{Bin}(n,p)$, and let $\lambda_n \coloneqq \mathbb{E}(X_n)$ be the mean of X_n. Let n_k be a sequence of positive integers satisfying $n_k = \left(1 + o\left(\frac{1}{\omega(n)}\right)\right) n$ for some function $\omega(n) \to \infty$. Let $\{Y_k\}_{k\in\mathbb{N}}$ be a sequence of random variables with distribution $Y_k \sim \text{Bin}(n_k, p)$ for all $k \in \mathbb{N}$. If $t = t(n)$ (with $t(n) > 0$ for sufficiently large n) satisfies $t = \Omega(\lambda)$ then we have that*

$$\mathbb{P}(X_k \leq \mathbb{E}(X_k) - t) \leq \exp\left(-\frac{t^2}{2\lambda}\left(1 + O_k\left(\frac{1}{\omega(n)}\right)\right)\right) = \exp\left(-\frac{t^2}{2\lambda}\left(1 + o_k(1)\right)\right), \tag{2.10}$$

where we write O_k and o_k to indicate that the inequality is in the limit as $k \to \infty$.

Proof. Let $\{Y_k\}_{k\in\mathbb{N}}$ and $\{X_n\}_n \in \mathbb{N}$ be as in the statement of the theorem, and let $t = t(n)$ be such that $t = \Omega(n)$. Then since $n_k = \left(1 + o\left(\frac{1}{\omega(n)}\right)\right) n$, it follows that for any $k \in \mathbb{N}$,

$$\mathbb{E}(X_k) = n_k p = \left(1 + o\left(\frac{1}{\omega(n)}\right)\right) np = \left(1 + o\left(\frac{1}{\omega(n)}\right)\right) \lambda.$$

This gives us that

$$\mathbb{P}(X_k \leq \lambda - t) = \mathbb{P}\left(X_k \leq \lambda\left(\left(1 + o\left(\frac{1}{\omega(n)}\right)\right) - \left(\left(t + o\left(\frac{\lambda}{\omega(n)}\right)\right)\right)\right). \tag{2.11}$$

Hence for any fixed $k \in \mathbb{N}$, we can apply Chernoff's Bound (Theorem 2.39) to (2.11), which gives

$$\mathbb{P}(X_k \leq \lambda - t) \leq \exp\left(-\frac{\left(t + o\left(\frac{\lambda}{\omega(n)}\right)\right)^2}{2\lambda\left(1 + o\left(\frac{1}{\omega(n)}\right)\right)}\right). \tag{2.12}$$

We focus on the expression inside the exponential function in (2.12). By (2.6) from Section 2.6, we have that

$$\frac{1}{1+o\left(\frac{1}{\omega(n)}\right)} = 1 + O\left(\frac{1}{\omega(n)}\right).$$

Hence we can write

$$\frac{\left(t+o\left(\frac{\lambda}{\omega(n)}\right)\right)^2}{2\lambda\left(1+o\left(\frac{1}{\omega(n)}\right)\right)} = \frac{\left(t+o\left(\frac{\lambda}{\omega(n)}\right)\right)^2}{2\lambda}\left(1+O\left(\frac{1}{\omega(n)}\right)\right). \tag{2.13}$$

Moreover, we can rewrite the numerator of (2.12) as

$$\left(t+o\left(\frac{\lambda}{\omega(n)}\right)\right)^2 = t^2\left(1+o\left(\frac{\lambda}{t\omega(n)}\right)\right)^2.$$

Then since $t = \Omega(\lambda)$, it follows that $\frac{\lambda}{t\omega(n)} = O\left(\frac{1}{\omega(n)}\right)$, and therefore

$$t^2\left(1+o\left(\frac{\lambda}{t\omega(n)}\right)\right)^2 = t^2\left(1+o\left(\frac{1}{\omega(n)}\right)\right)^2 = t^2\left(1+o\left(\frac{1}{\omega(n)}\right)\right).$$

Putting this together with (2.13), we conclude that

$$\begin{aligned}
\mathbb{P}(X_k \leq \lambda - t) &\leq \exp\left(-\frac{\left(t+o\left(\frac{\lambda}{\omega(n)}\right)\right)^2}{2\lambda\left(1+o\left(\frac{1}{\omega(n)}\right)\right)}\right), \\
&= \exp\left(-\frac{t^2}{2\lambda}\left(1+o\left(\frac{1}{\omega(n)}\right)\right)\left(1+O\left(\frac{1}{\omega(n)}\right)\right)\right) \\
&= \exp\left(-\frac{t^2}{2\lambda}\left(1+O\left(\frac{1}{\omega(n)}\right)\right)\right),
\end{aligned}$$

completing the proof. □

Remark 2.42. *We note that although we required $t > 0$ in our proof of Theorem 2.41, the theorem still holds for $t = 0$. In this case the inequality is trivial.*

Theorem 2.41 says that given a sequence of random variables converging to a binomial distribution we can apply Chernoff's bound with the asymptotic mean of the

sequence. In particular, if $\frac{t^2}{2\lambda} \to \infty$, then

$$\exp\left(-\frac{t^2}{2\lambda}(1+o(1))\right) = \frac{\exp\left(o\left(\frac{t^2}{2\lambda}\right)\right)}{\exp\left(\frac{t^2}{2\lambda}\right)} \to 0.$$

So in this case, $\mathbb{P}(X_k \leq \mathbb{E}(X_k) - t) \to 0$. Note that since $t = \Omega(\lambda)$, a sufficient condition for $t^2/(2\lambda) \to \infty$ is that $\lambda \to \infty$. Unfortunately we will have λ equal to a constant in our application of Theorem 2.41.

This chapter has described the necessary theoretical background for the thesis. In the chapter that follows, we will use these theoretical tools to study the existence of a specific graph property in the binomial random graph $\mathcal{G}_{n,p}$.

Chapter 3

The Giant Component in Erdős-Rényi Random Graphs

In the mid 19th century certain Victorian aristocrats were becoming concerned that their surnames might die out in the coming generations. Polymath (and half-cousin of Charles Darwin) Sir Francis Galton posed the following question in the 1873 Educational Times:

How many male children (on average) must each generation of a family have in order for the family name to continue in perpetuity?

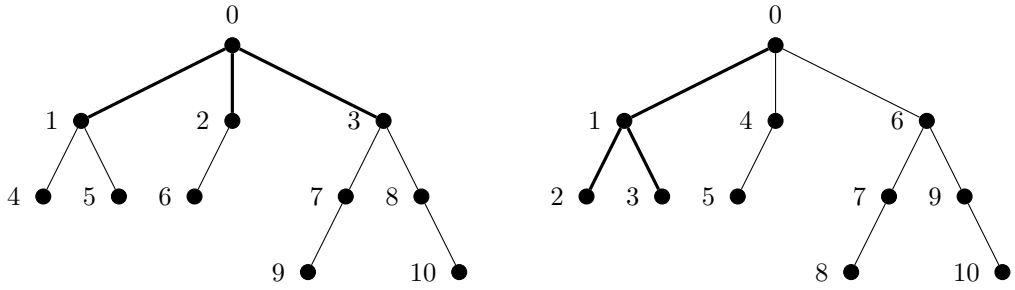
It was Reverend Henry William Watson who provide an answer, and in 1874 the Galton & Watson wrote a paper on the matter entitled *On the probability of extinction of families* [54].

In this chapter we study the threshold for the emergence of the giant component in binomial random graphs. A giant component in a graph is a connected set of vertices which contains a constant fraction of all of the vertices. The property of having a giant component is a monotone increasing property of graphs, and as such, exhibits a threshold by Theorem 2.35. We prove that $p = \frac{1}{n}$ is a sharp threshold for the emergence of a giant component (Theorems 3.10 and 3.11), and that when a giant component exists, it is unique (Lemma 3.13). Hence we routinely refer to this component as *the* giant component. The phase transition which occurs in $\mathcal{G}_{n,p}$ around $p = \frac{1}{n}$ is known as the *evolution* of the giant component. We note here that in the physics literature, which we look at in Chapter 4, the study of connected components of a random graph is called *percolation theory*.

The approach we follow makes use of the technique developed by Galton & Watson [54] known as a *branching process*. Though these processes were discovered (and

studied) in the late 19th century, the diversity of their applicability was not realised until the latter half of the 20th century [3]. Since then, they have become an extremely important tool in the theory of random graphs, and we will continue to use them in Chapter 4. Branching processes allow us to approximate the structure of a random graph, and under certain conditions, this approximation turns out to be very precise asymptotically. We will need to understand some basic results in the theory of branching processes before we are able to apply them to random graphs.

The branching process we consider is also known as a *breadth first search*, as opposed to a *depth first search*. The difference is illustrated in the image below. A breadth first search is depicted on the left side of the image, and a depth first search on the right, both beginning at vertex 0. The order of the search is indicated by the numbers $0, 1, \ldots, 10$, and we have used thick lines to highlight the difference between the two algorithms in the first three stages of searching. Krivelevich & Sudakov (2012) [34] prove theorems analogous to Theorems 3.10 and 3.11 using depth first searches, and were the first to do so.



3.1 Branching Processes

We begin this section by introducing the model proposed by Galton & Watson [54].

3.1.1 Uniform (Galton-Watson) Branching Process

Let $\{Z_{n,k}\}_{n\in\mathbb{N},\ k\in\mathbb{Z}^+}$ be a sequence of independent and identically distributed random variables with distribution $\mathcal{D}$, and taking values in the non-negative integers. We call each n a *generation*, and we call $\mathcal{D}$ the *offspring distribution*. Furthermore, for $Z \sim \mathcal{D}$ we write the probability mass function (p.m.f.) for Z as $\{p_i\}_{i\in\mathbb{N}}$, where $p_i = \mathbb{P}(Z = i)$.

(I) A population starts with one individual at time $n = 0 : Z_0 = 1$.

(II) After one unit of time ($n = 1$), the individual Z_0 produces $Z_1 := Z_{0,1}$ identical copies of itself and dies, where $Z_1 \sim \mathcal{D}$.

(III) (a) If $Z_1 = 0$, the population dies and $Z_n = 0$ for all $n \geq 1$.

(b) If $Z_1 > 0$, then at time $n = 2$, each of the Z_1 individuals give birth to a random number of children and dies. The first individual has $Z_{1,1}$ children, the second has $Z_{1,2}$, ..., and the last one has Z_{1,Z_1} children. Let $Z_{n,i}$ denote the number of children of the i-th individual in the n-th generation, then each $Z_{n,i}$ is independently distributed according to $\mathcal{D}$. The total number of individuals in the second generation is then

$$Z_2 = \sum_{k=1}^{Z_1} Z_{1,k}.$$

(c) The third, fourth, etc. generations are produced in the same way. That is,

$$Z_{n+1} = \sum_{k=1}^{Z_n} Z_{n,k}.$$

If it happens that $Z_n = 0$ for some $n \in \mathbb{N}$, then $Z_m = 0$ for all $m \geq n$; if this occurs we say that the population is *extinct*.

Definition 3.1. A sequence $\{Z_n\}_{n\in\mathbb{N}}$ with the properties described in I, II, and III above is called a *uniform* (or Galton-Watson) *branching process*.

3.1.2 Generating Functions and Uniform Branching Processes

The main point of interest in analyzing the *uniform branching process* is looking at the probabilistic properties of the sequence $\{Z_n\}_{n\in\mathbb{N}}$. Remember that Z_n is the total number of individuals in the n-th generation. We have that by definition, Z_n is the sum of Z_{n-1} independent copies of a random variable with the offspring distribution. The distribution of Z_n is completely determined by its p.m.f., which we have already seen in Remark 2.19 is completely determined by its probability generating function. Indeed while we may not always be able to compute the p.m.f. it is often possible to compute the generating function.

Proposition 3.2. *Let $\{Z_n\}_{n\in\mathbb{N}}$ be a branching process, and let the generating function of its offspring distribution $\{p_n\}_{n\in\mathbb{N}}$ be given by $f(z)$. Then the generating*

function of Z_n is the n-fold composition of f with itself, that is,

$$f_{Z_n}(z) = \underbrace{f(f(\dots f}_{n}(z)\dots)) = f^n(z), \quad n \geq 1.$$

Proof. We proceed by induction. When $n = 1$, we have that the distribution of Z_1 is exactly the offspring distribution, hence $f_{Z_1} = f$. Now suppose the proposition is true for some $n = k$, and consider the case when $n = k + 1$. We have

$$Z_{k+1} = \sum_{i=1}^{Z_k} Z_{k,i}$$

is the sum of Z_k independent random variables with the offspring distribution $\{p_n\}_{n\in\mathbb{N}}$. But by Theorem 2.22, we have that

$$f_{Z_{k+1}}(z) = f_{Z_k}(f(z)) = f^k(f(z)) = f^{k+1}(z),$$

where the second equality comes from the inductive assumption. Hence the proposition holds for $n = k + 1$ and thus for all $n \in \mathbb{N}$ by induction. □

3.1.3 Extinction Probability

The original question which Galton & Watson sought to answer can be stated as follows:

Under what conditions on the offspring distribution will the process $\{Z_n\}_{n\in\mathbb{N}}$ never go extinct, that is, when does

$$\mathbb{P}(Z_n \geq 1 \text{ for all } n \in \mathbb{N}) = 1$$

hold?

We can provide an answer to this question using generating functions. Let $Z = \sum_{n\geq 0} Z_n$ be the total number of offspring in the branching process. The probability ρ of extinction of the branching process is defined to be

$$\rho = \mathbb{P}(Z < \infty) = \lim_{n\to\infty} \mathbb{P}(Z_n = 0),$$

since if $Z_n = 0$ for some $n \in \mathbb{N}$, then $Z_m = 0$ for all $m \geq n$. Then by Lemma 2.20, we have that $\mathbb{P}(Z_n = 0) = f_{Z_n}(0)$, and so by Proposition 3.2,

$$\rho = \lim_{n\to\infty} \mathbb{P}(Z_n = 0) = \lim_{n\to\infty} f_{Z_n}(0) = \lim_{n\to\infty} f^n(0).$$

We may now state and prove the following important theorem.

Theorem 3.3. *The extinction probability ρ is the smallest non-negative solution of the equation*

$$z = f(z),$$

where f is the generating function of the offspring distribution.

Proof. We first prove that ρ is a solution to the equation $z = f(z)$. Since f is continuous inside its radius of convergence (see Lemma 2.20), we have that for any convergent sequence $\{\rho_n\}_{n\in\mathbb{N}}$, with $\rho_n \in [0, 1]$ for all n,

$$f\left(\lim_{n\to\infty} \rho_n\right) = \lim_{n\to\infty} f(\rho_n).$$

Consider then the sequence

$$\rho_n = f^n(0).$$

One has that $\rho_n \to \rho$ as $n \to \infty$, and that $f(\rho_n) = \rho_{n+1}$. Therefore

$$\rho = \lim_{n\to\infty} \rho_n = \lim_{n\to\infty} \rho_{n+1} = \lim_{n\to\infty} f(\rho_n) = f\left(\lim_{n\to\infty} \rho_n\right) = f(\rho),$$

and so ρ solves the equation $f(z) = z$.

We now show that ρ is indeed the smallest such non-negative solution. Suppose that ρ' is another solution to $f(z) = z$ on $[0, 1]$. Then since $0 \leq \rho'$, and f is a non-decreasing function (again, see Lemma 2.20), we have that

$$f(0) \leq f(\rho') = \rho'.$$

Then applying the generating function to both sides, we have that

$$f(f(0)) \leq f(f(\rho')) = f(\rho') = \rho'.$$

Continuing in this way we have that for any $n \in \mathbb{N}$,

$$f^n(0) \leq f^n(\rho') = \rho'.$$

It follows that $\rho = \lim_{n\to\infty} f^n(0) \leq \lim_{n\to\infty} \rho' = \rho'$, and so ρ' does not exceed ρ. This completes the proof. □

Corollary 3.4. *Keep the above notation and let Z be distributed according to the offspring distribution $\{p_n\}_{n\in\mathbb{N}}$ where $p_n = \mathbb{P}(Z = n)$. If $p_1 \neq 1$ then $\rho = 1$ if and only if $\mathbb{E}(Z) \leq 1$. If $p_1 = 1$ then clearly $\rho = 0$.*

Proof. If $p_0 = 0$, then $\rho = 0$ and $\mathbb{E}(Z) > 1$, so suppose $p_0 > 0$. Suppose that $p_1 \neq 1$. Recall from Lemma 2.20 that $f'(1) = \mathbb{E}(Z)$, and since $p_1 \neq 1$ by assumption, it is not true that $f(z) = z$ everywhere. Consider then the two curves $y = f(z)$ and $y = z$. If $f'(1) < 1$, then by the fact that f is continuous, nondecreasing, and convex on $[0, 1]$ (Lemma 2.20), the curves intersects at a single point. This point cannot be 0 since we assumed $p_0 > 0$, and hence since again by Lemma 2.20, $f(1) = 1$ we see that it must be at $z = 1$. Therefore by Theorem 3.3, $\rho = 1$. If instead $f'(1) > 1$, then there are two intersections between the curves. Clearly $z = 1$ is still a point of intersection, but the second is at some $z < 1$, and Theorem 3.3 tells us that that this is precisely where $z = \rho$. Hence $\rho = 1$ if and only if $\mathbb{E}(Z) \leq 1$. □

We now consider two key examples of extinction probabilities.

Example 3.5. *Suppose that the number of offspring is Poisson with distribution $X \sim \mathrm{Po}(\lambda)$. Then the corresponding generating function is*

$$f_X(z) = \sum_{k=0}^{\infty} \frac{c^k z^k}{k!} e^{-\lambda} = \exp(\lambda(z-1)).$$

Thus if $\lambda > 1$, then by Corollary 3.4, the extinction probability ρ is equal to $1-\beta(\lambda)$, where $\beta = \beta(\lambda) \in (0, 1)$ is uniquely determined by the largest non-negative solution to equation

$$\beta + e^{-\beta\lambda} = 1. \tag{3.1}$$

This can be seen by putting $z = 1-\beta$ into the fixed point equation $z = \exp(\lambda(z-1))$.

Example 3.6. *Suppose that the number of offspring is Binomial with distribution* $X_n \sim \text{Bin}(n,p)$*, where* $np \to \lambda > 1$ *as* $n \to \infty$*. Then the corresponding generating function is*

$$f_{X_n}(z) = \sum_{k=0}^{n} \binom{n}{k} (zp)^k (1-p)^{n-k} = (zp + 1 - p)^n,$$

for every real number z *we have*

$$\lim_{n\to\infty} f_{X_n}(z) = \exp(\lambda(z-1)) = f_X(x).$$

That is, the p.g.f. of X_n *tends pointwise to the p.g.f. of* $X \sim \text{Po}(\lambda)$*. Hence as* $n \to \infty$*, the probability of extinction* $\rho_{n,c}$ *of the branching process defined by* X_n *converges to* $1 - \beta(\lambda)$*, where* $\beta(\lambda)$ *is defined in (3.1).*

Remark 3.7. *We can make a few minor adjustments to generalise the branching process proposed by Galton & Watson. We again let* $\{Z_{n,k}\}_{n\in\mathbb{N},k\in\mathbb{Z}^+}$ *be a sequence of independent random variables, but now allow* $Z_{n,k}$ *to have p.m.f.* $\{p_i(n,k)\}_{i\in\mathbb{N}}$*. In this case, the offspring distribution does not need to be the same for every individual, and we call this a non-uniform branching process.*

We now discuss an alternative description of the non-uniform branching process.

3.1.4 Exploration of Branching Processes

It is useful to consider a branching process unfolding step by step rather than generation by generation; that is, indexing the branching process in $\mathbb{N}$ rather than $\mathbb{N} \times \mathbb{N}$. As described in Section 3.1.5, when applying the branching process to the exploration of a random graph, we can imagine that the branching process has already unfolded and we are now deducing what is most likely to have happened, vertex by vertex (each vertex represented by a single subscript). In light of this, we present an original characterisation of non-uniform branching processes.

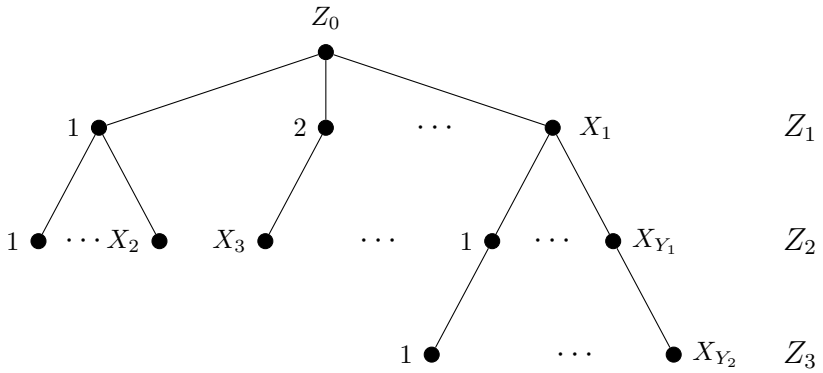
Formally, let $X_1, X_2, \ldots$ be a sequence of independent non-negative integer-valued random variables. Then one can equivalently define a *non-uniform branching process* (as in Remark 3.7) with underlying space $X_1, \ldots, X_n$ as follows. At time $n = 0$ we begin with a single particle. At time $n = 1$, this particle gives birth to $X_1 \geq 0$ other particles and then dies. Each of those X_1 particles then gives birth to $X_2, \ldots, X_{X_1+1}$ more particles and each die also. This process continues either infinitely or until all particles have died. We call the initial X_1 children the *first*

generation, the following $X_2, \ldots, X_{X_1+1}$ the *second generation* and so on. Let Z_i denote the number of offspring in the i-th generation, and let $Y_i = \sum_{j=0}^{i} Z_j$. Then we define $Z_0 = 1$ (with probability 1), while for $i \geq 1$, the variable Z_i is given by

$$Z_i = \sum_{j=Y_{i-2}+1}^{Y_{i-1}} X_j. \tag{3.2}$$

Note that if $Z_i = 0$ for some i, then $Z_j = 0$ for all $j \geq i$. These Z_i are exactly the description of the branching process we are used to from Section 3.1.1, but now the summation that specifies them is indexed by only 1 subscript. We use i instead of n here as the subscript because when we apply this to a random graph, n will denote the number of vertices in the graph.

In practice we avoid the messy expression (3.2) above by using a uniform branching process to provide an upper and lower bound (in the sense of stochastic dominance) for the non-uniform process. Indeed, we will only perform calculations with the case where $X_i = X$ for every $i = 1, \ldots, n$, which is simply the *uniform branching process* we saw earlier. As we have already remarked, in this case we have that Z_i is the sum of Z_{i-1} independent copies of X. See below for a visual representation of this process.



We may now give a third formulation of the branching process in terms of a recurrence relation. The following formulation can be found in [41]. Given i.i.d. variables $X_1, X_2, \ldots$ as above with offspring distribution X, a branching process on $X_1, X_2, \ldots$ can equivalently be described as a sequence of random variables $\{G_i\}_{i \in \mathbb{N}}$ satisfying the following recurrence relation

$$G_0 = 1, \qquad G_i = G_{i-1} + X_i - 1 \quad (i > 0),$$

which gives explicit formula

$$G_i = 1 - i + \sum_{j=1}^{i} X_j \qquad \text{for all } i \geq 0.$$

We will refer to this as a *branching recurrence process* (BRP) on $X_1, X_2, \ldots$, in order to differentiate it from the way we have formalised branching processes so far. We call G_i the number of *alive* particles at step i. If $G_i = 0$ for some $i > 0$, we say that the process has gone *extinct*. Let T be the first index for which $G_T = 0$, and let $T = \infty$ if extinction does not happen. Then we have the following inequality.: If $1 \leq i \leq T$, then $G_i \geq 0$ and hence

$$\sum_{j=1}^{i} X_j \geq i - 1. \tag{3.3}$$

Note that if the process dies at time T, exactly T particles were active during the process. Hence we can call T the *total population*.

Finally then, we can present the application of branching processes to our question in the introduction regarding the evolution of the giant component in $\mathcal{G}_{n,p}$.

3.1.5 Branching Processes on $\mathcal{G}_{n,p}$

The branching process provides us with a useful framework for analysing the size of components in the random graphs $\mathcal{G}_{n,p}$ or $\mathcal{G}_{n,M}$. Consider the binomial random graph $\mathcal{G}_{n,p}$ and choose a vertex v_1 from it at random. We call v_1 the *root* of the branching process. Then v_1 has $X_1 \sim \text{Bin}(n-1, p)$ neighbours: label them as $v_2, v_3, \ldots, v_{X_1+1}$. We can view v_1 as a particle that gives birth to these X_1 new particles and dies. Now v_2 can have up to $n-(X_1+1)$ "non-discovered" neighbours, each with probability p. Let $X_2 \sim \text{Bin}(n-1-X_1, p)$ be the number of "non-discovered" neighbours of v_2, and label them as $v_{X_1+2}, v_{X_1+3}, \ldots, v_{X_1+X_2+1}$. Continuing in this way, we can construct a sequence of random variables $X_1, X_2, \ldots$ taking on non-negative integer-values and this sequence is a candidate from which we can define a BRP.

Let $\{G_i\}_{i \in \mathbb{N}}$ be the BRP on $X_1, X_2, \ldots$. Notice firstly that this is certainly not a uniform branching process, since each particle may have a different offspring

distribution. Indeed, we have

$$G_i = 1 + \sum_{j=1}^{i} X_j - i \quad (i \geq 0) \tag{3.4}$$

Then

$$X_i \sim \text{Bin}(n - 1 - \sum_{j=1}^{i-1} X_j, p)$$

$$\text{which implies that } X_i \sim \text{Bin}(n - (i-1) - G_{i-1}, p). \tag{3.5}$$

For a BRP on a random graph, we will say that a vertex v is *saturated* if all of its neighbours have been found, and *unsaturated* if it has been discovered in the branching process but not all of its neighbours have yet been found.

Intuitively, (3.5) says that the possible neighbours of v_i in the branching process can be any vertex which has not already had children (there are $n - (i-1)$ of these), or has not already been a child of another vertex. The vertices which have already been the child of another vertex are the G_{i-1} "alive" vertices. Moreover T, which represents the total population of the branching process, is the size of the connected component containing v_1. Hence studying T gives insight into the size of components in $\mathcal{G}_{n,p}$.

Remark 3.8. *As we mentioned earlier, we prefer not to deal with non-uniform processes. Performing calculations with non-uniform processes can quickly become complicated, and often does not yield a closed form solution. Hence we will bound the BRP on* $X_i \sim \text{Bin}(n-(i-1)-G_{i-1}, p)$ *above and below as follows: Let* $\{G_i^+\}_{i\in\mathbb{N}}$ *and* $\{G_i^-\}_{i\in\mathbb{N}}$ *be BRP's with corresponding offspring distributions*

$$X^+ \sim \text{Bin}(n, p), \quad X^- \sim \text{Bin}(n - \lambda k, p)$$

for some constant $\lambda > 0$*, and an integer-valued function* $k = k(n)$ *satisfying* $0 < \lambda k < n$*, with* $k(n) = o(n)$ *for all* $n \in \mathbb{N}$*. Then (as in Example 2.26) we have that in terms of stochastic dominance* $X_i^- \preceq X_i \preceq X_i^+$*, that is,*

$$\mathbb{P}(X_i^+ \geq x) \geq \mathbb{P}(X_i \geq x) \geq \mathbb{P}(X_i^- \geq x), \quad \textit{for all } x \in \mathbb{N}.$$

The inequalities above are used implicitly in the proof of Theorems 3.10 and 3.11 found in [32, Chapter 5]. We use these inequalities to provide an ordering of BRPs on binomial random variables in terms of stochastic dominance. To the best of our knowledge, this has not been done explicitly before.

Corollary 3.9 (Original)**.** *Let $\{X_j\}_{j\in\mathbb{N}}$ be a sequence of random variables with $X_j \sim \text{Bin}(n_j, p)$ for positive integers n_j, where $j \in \mathbb{N}$. Similarly, let $\{Y_j\}_{j\in\mathbb{N}}$ be sequences of random variables $Y_j \sim \text{Bin}(m_j, p)$ for positive integers m_j, where $j \in \mathbb{N}$. Moreover, let $\{G_i\}_{i\in\mathbb{N}}$ be a BRP on $\{X_i\}_{i\in\mathbb{N}}$, and $\{G'_i\}_{i\in\mathbb{N}}$ a BRP on $\{Y_i\}_{i=1}^{\infty}$. If $n_j < m_j$ for all $j \in \mathbb{N}$ then $G_i \preceq G'_i$ for all $i \in \mathbb{N}$.*

Proof. Let $\{X_i\}_{i\in\mathbb{N}}$, $\{Y_i\}_{i\in\mathbb{N}}$, $\{G_i\}_{i\in\mathbb{N}}$ and $\{G'_i\}_{i\in\mathbb{N}}$ be as described in the statement of the corollary. Then

$$\mathbb{P}(G_i = k) = \mathbb{P}\bigg(\sum_{j=1}^{i} X_j = k + i - 1\bigg),$$
$$\mathbb{P}(G'_i = k) = \mathbb{P}\bigg(\sum_{j=1}^{i} Y_j = k + i - 1\bigg).$$

By Lemma 2.27,

$$\sum_{j=1}^{i} X_j \sim \text{Bin}\bigg(\sum_{j=1}^{m} n_j, p\bigg) \quad \text{and} \quad \sum_{j=1}^{i} Y_j \sim \text{Bin}\bigg(\sum_{j=1}^{m} m_j, p\bigg).$$

But we already know from Example 2.26 that if $m < n$ then $\text{Bin}(m, p) \preceq \text{Bin}(n, p)$. Hence if $m_j > n_j$ for all $j \in \mathbb{N}$, then

$$\sum_{j=1}^{i} X_j \preceq \sum_{j=1}^{i} Y_j$$

holds for all $i \in \mathbb{N}$. The result follows immediately. □

3.2 Threshold for existence of the Giant Component

Recall that a giant component in a graph is a connected set of vertices which contains a constant fraction of all of the vertices. We stated in the introduction to this chapter that $p = \frac{1}{n}$ is a sharp threshold for the emergence of the giant component, and we will prove this result shortly. The values of p for which $p = \frac{1-\epsilon}{n}$

for some $\epsilon > 0$ are said to be in the *subcritical* regime. The values of p for which $p = \frac{1+\epsilon}{n}$ for some $\epsilon > 0$ are said to be in the *supercritical* regime. When $p = \frac{1+o(1)}{n}$, this is called the *critical window*. Analysis of the giant component inside the critical window is rather delicate and is beyond the scope of this thesis, though it was proved by Łuczak [38] that when $p = \frac{1}{n} + \frac{\lambda}{n^{4/3}}$ for a constant λ, the largest component of $\mathcal{G}_{n,p}$ is on the order of $\Theta(n^{2/3})$. We now present an important theorem of Erdős and Rènyi (1960).

Theorem 3.10 (Subcritical case)**.** *Let $p = \frac{\lambda}{n}$, where $\lambda < 1$ is a constant. Then w.h.p. the largest component of $\mathcal{G}_{n,p}$ has at most $k(n) = \frac{2+\epsilon}{(1-\lambda)^2} \log n$ vertices, where $\epsilon > 0$.*

Proof. Let us assume that $p = \frac{\lambda}{n}$ and $\lambda < 1$. Our goal here is to show that the probability that an arbitrary vertex v is in a component of size greater than $k = k(n)$ goes to 0 as $n \to \infty$. Hence we choose a vertex v of $\mathcal{G}_{n,p}$ and take the BRP as in Section 3.1.5 to explore the component containing v. We have that the probability that v is in a component of size at least k is exactly the probability that the BRP has not gone extinct before k rounds of exploration. Hence

$$\mathbb{P}(v \text{ is in component of size at least } k) = \mathbb{P}(k \leq T) \leq \mathbb{P}\left(\sum_{j=1}^{k} X_j \geq k-1\right)$$

by (3.3). Moreover, by Corollary 3.9, we see that

$$\mathbb{P}\left(\sum_{j=1}^{k} X_j \geq k-1\right) \leq \mathbb{P}\left(\sum_{j=1}^{k} X_j^+ \geq k-1\right).$$

Finally then, since we had n choices for our initial vertex v, the probability that $\mathcal{G}_{n,p}$ contains a component of size at least $k \geq (2+\epsilon)\log n/(1-\lambda)^2$ is bounded above by

$$\begin{aligned} n\,\mathbb{P}\left(\sum_{j=1}^{k} X_j^+ \geq k-1\right) &= n\mathbb{P}\left(\sum_{j=1}^{k} X_j^+ \geq k\lambda + (1-\lambda)k - 1\right) \\ &\leq n\exp\left(-\frac{((1-\lambda)k-1)^2}{2(\lambda k + (1-\lambda)k/3)}\right) \\ &\leq n\exp\left(-\frac{(1-\lambda)^2}{2}k\right) \end{aligned} \tag{3.6}$$

$$\leq n^{-\epsilon/2}$$
$$= o(1).$$

Here the first inequality comes from applying Chernoff's bound (see Theorem 2.39) with $\mathbb{E}\big(\sum_{j=1}^{k} X_j^+\big) = knp = k\lambda$. □

Theorem 3.10 proves that when $p = \frac{1-\epsilon}{n}$ for some $\epsilon > 0$, no giant component exists. To prove that a unique giant component exists for $p > \frac{1+\epsilon}{n}$, we need to take a little more care. We show first that there are no components of "medium" size, and then that there is at most one component of "large" size. We finish the proof by showing exactly how many vertices are in components of "small" size and "large" size. We divide the proof of Theorem 3.11 into three lemmas.

Theorem 3.11 (Supercritical case)**.** *Let $p = \frac{\lambda}{n}$, where $\lambda > 1$ is a constant. Let $\beta = \beta(\lambda) \in (0, 1)$ be the unique smallest positive solution to equation*

$$\beta + e^{-\beta\lambda} = 1$$

as in Example 3.5. Then $\mathcal{G}_{n,p}$ contains a giant component of $(1 + o_p(1))\beta n$ vertices. Furthermore, w.h.p. the size of the second largest component of $\mathcal{G}_{n,p}$ is at most $\frac{(40+\epsilon)\lambda}{3(\lambda-1)^2} \log n$ for any $\epsilon > 0$.

Going back to our definitions in probability asymptotics from Section 2.3, a reformulation of the first part of this theorem is that for every $\epsilon > 0$, w.h.p. $\mathcal{G}_{n,p}$ contains a component of size $(1 + \epsilon)\,\beta_n$. In Figure 3.1 below, we plot the function $\beta(\lambda)$ against λ, where $\lambda > 1$ and $\beta(\lambda)$ is defined in Theorem 3.11. One can see the rapid increase in the size of the giant component as λ gets larger. We now prove the first of three lemmas. Suppose for the remainder of this chapter that $\lambda > 1$ is a constant, and $p = \frac{\lambda}{n}$. Lemma 3.12 makes use of the branching processes $\{G_i^-\}_{i\in\mathbb{N}}$ and $\{G_i^+\}_{i\in\mathbb{N}}$ defined in Remark 3.8.

Lemma 3.12. *Let $k_- = \frac{(40+\epsilon)\lambda}{3(\lambda-1)^2} \log n$ for some $\epsilon > 0$ and $k_+ = n^{2/3}$. Then w.h.p. there is no component of $\mathcal{G}_{n,p}$ with size in the interval $[k_-, k_+]$.*

Proof. Let v be a vertex in $\mathcal{G}_{n,p}$ and let k be such that $k_- \leq k \leq k_+$. We consider a BRP beginning at v. In order to show that there is no component of $\mathcal{G}_{n,p}$ with size

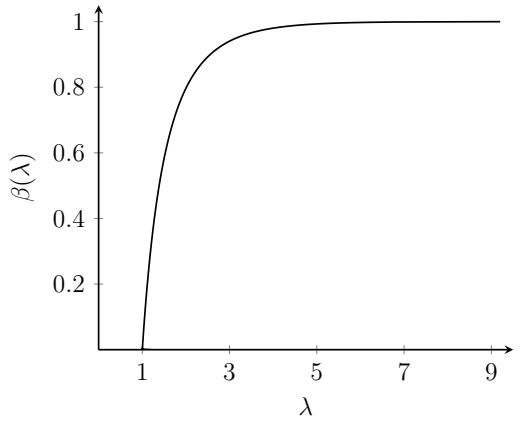


Figure 3.1: The fraction $\beta(\lambda)$ of vertices in the giant component

k, it suffices to show that at the k-th step there is always some positive number of alive vertices remaining in the BRP. In particular, we show that at the k-th step there are at least $(\lambda-1)k/2$ alive vertices in the component containing v, that is, $G_k \geq (\lambda-1)k/2$. To do this, we show that probability of the complementary event vanishes, that is,

$$\mathbb{P}\left(\text{there exists } k_- \leq k \leq k_+ \text{ s.t. } G_k < \frac{(\lambda-1)k}{2}\right) \tag{3.7}$$

tends to 0. Why is (3.7) true? Well, suppose such a k exists with $k_- \leq k \leq k_+$ and $G_k < \frac{(\lambda-1)k}{2}$. Then notice that for any $i \leq k$,

$$i + G_i \leq k + G_k < \frac{(\lambda+1)k}{2} < \lambda k \leq \lambda k_+.$$

Recalling that for any $i = 1, \ldots, k_+$, each variable X_i in the BRP on $\{X_i\}_{i\in\mathbb{N}}$ is distributed as $X_{i+1} \sim \text{Bin}(n-(i+G_i), p)$ $(i \geq 0)$, each X_i stochastically dominates

$$X_i^- \sim \text{Bin}(n - \lambda k_+, p)$$

as in Remark 3.8. To prepare to apply Chernoff's bound, we note that for the distribution X_i^- we have

$$\mathbb{E}\left(\sum_{i=1}^{k} X_i^-\right) = \lambda k(1-\lambda n^{-1/3}) = \lambda k\left(1-o(1)\right).$$

Since X_i^- and X_i^+ are binomially distributed random variables, and $\mathbb{E}(X_i^-) \to \mathbb{E}(X_i^+)$, Theorem 2.41 allows us to apply Chernoff's bound to $\sum_{i=1}^{k} X_i^-$ using the

asymptotic mean of the sequence $\{X_i^-\}$, that is, using λk. Note that $\frac{2+(\lambda-1)k}{2} = \Omega(\lambda k)$ and so the condition "$t = \Omega(\lambda)$" required by Theorem 2.41 is satisfied when we use this theorem in (3.8). Hence first applying the union bound (see Lemma 2.17), followed by the asymptotic version of Chernoff's bound (Theorem 2.41), we have that for any specific k where $k^- \leq k \leq k^+$,

$$\begin{aligned}
\mathbb{P}\left(G_k < \frac{(\lambda-1)k}{2}\right) &= \mathbb{P}\left(\sum_{i=1}^{k} X_i < k - 1 + \frac{(\lambda-1)k}{2}\right) \\
&\leq \mathbb{P}\left(\sum_{i=1}^{k} X_i^- < \lambda k - \frac{2+(\lambda-1)k}{2}\right) \\
&\leq \exp\left(-\frac{(2+(\lambda-1)k)^2}{8\lambda k}(1+o(1))\right) \qquad (3.8) \\
&= \exp\left(-\frac{4+4(\lambda-1)k+(\lambda-1)^2}{8\lambda k}(1+o(1))\right) \\
&= \exp\left(-\frac{(\lambda-1)^2 k}{8\lambda}(1+o(1))\right)
\end{aligned}$$

Finally, summing over all $k \in \{k^-, \ldots, k^+\}$ and multiplying by n for the initial choice of vertex v, we have that (3.7) is bounded above by

$$\begin{aligned}
n\sum_{k=k_-}^{k_+} \exp\left(-\frac{(\lambda-1)^2 k}{8\lambda}\right) &\leq n\sum_{k=k_-}^{k_+} \exp\left(-\frac{(\lambda-1)^2 k}{8\lambda}(1+o(1))\right) \\
&\leq nk_+ \exp\left(-\frac{(\lambda-1)^2 k_-}{8\lambda}(1+o(1))\right) \\
&= n^{5/3}\exp\left(-\frac{40+\epsilon}{24}\log n\right) \\
&= n^{-\epsilon(1+o(1))/24-o(1)} \\
&\leq n^{-\epsilon/25} \\
&= o(1),
\end{aligned}$$

completing the proof. □

Lemma 3.13. *With high probability there is at most one component of size at least k_+.*

Proof. Consider a pair of vertices $v, v' \in V(\mathcal{G}_{n,p})$, and suppose that v, v' belong to components of size at least k_+. First, we look at a BRP beginning at v. By Lemma 3.12, we know that after the first k_+ steps, there are at least $(\lambda-1)k_+/2$ alive vertices left in the exploration process. The same can be said if we start a separate BRP beginning at v'; in this process we either end up joining with the component from the BRP beginning at v, or we end up with a set of vertices separate from the component containing v, among which at least $(\lambda-1)k_+/2$ are still alive. If the BRP's beginning at v and v' join up, then we are done, so we want to bound the probability that they do not join up. That is, we are interested in the probability that there are no edges between the two sets of alive vertices of the two BRP's. If this probability vanishes then with w.h.p the two components are connected. Using (2.2) from Section 2.6, and noting that $n^2 = o(e^{n^{1/3}})$, this probability is bounded above by

$$\exp\left(\frac{((\lambda-1)k_+)^2}{4}\log(1-p)\right) \leq \exp\left(-(\lambda-1)^2\lambda n^{1/3}/4\right) = o(n^{-2}).$$

Multiplying by $n(n-1) = O(n^2)$ for the choice of v and v' gives that the probability that there are no edges between the two sets of alive vertices of two BRP's beginning at any two vertices of $\mathcal{G}_{n,p}$ is $o(1)$, completing the proof. □

Recall from Example 3.6 that the limiting extinction probability of a branching process with offspring distribution $\mathrm{Bin}(n,p)$ is given by $\rho = 1-\beta$, where β is the unique smallest positive solution of $\beta + e^{-\beta\lambda} = 1$. We may now present the final lemma that we need before proving Theorem 3.11 for the supercritical case.

Lemma 3.14. *Let X be the number of vertices of $\mathcal{G}_{n,p}$ in components of size at most k_-. Then $\mathbb{E}(X) = \rho n + o(n)$ and $\mathrm{Var}(X) = o(n^2)$, where ρ is the extinction probability of a uniform branching process with offspring distribution $\mathrm{Bin}(n,p)$.*

Proof. Let X be a random variable counting how many vertices of $\mathcal{G}_{n,p}$ are in components of size at most k_-. Then $X = \sum_v I_v$ where I_v is an indicator variable for the event "v is in a component of size at most k_-". We want to show that $E(I_v) = \rho + o(1)$, so we consider a BRP beginning at v. Since the distributions X^+ and X^- converge to $\mathrm{Bin}(n,p)$ as $n \to \infty$, we have that the extinction probability of a BRP on either X^+ or X^- is $\rho + o(1)$ (as $n \to \infty$). Recall that in Lemma 3.12

we showed that X_i^- is stochastically dominated by X_i for all $i \in \mathbb{N}$. It follows then from Corollary 3.9 that a BRP $\{G_i\}_{i\in\mathbb{N}}$ on X stochastically dominates the BRP on X^-; we will call this BRP G^-. Hence, if T is the total population of the BRP on X, we have

$$\mathbb{P}(G_T^- > 0) \leq \mathbb{P}(G_T > 0) = 0.$$

That is, the BRP on X^- becomes extinct and therefore $\mathbb{P}(I_v = 1) \leq \rho + o(1)$. Moreover, if X^+ goes extinct, then the probability that it does so before k_- iterations of the BRP is $1 - o(n^{-1})$. This is precisely what (3.6) states in the proof of Theorem 3.10. Now if X^+ goes extinct at time $i < k_-$, then $G_i \preceq G_i^+ = 0$, so v is in a component of size at most k_-. Hence $\rho + o(1) \leq \mathbb{P}(I_v = 1)$ and we conclude that $\mathbb{E}(I_v) = \mathbb{P}(I_v = 1) = \rho + o(1)$ as required.

We now compute the variance of X. Notice that

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}\left[\left(\sum_v I_v\right)^2\right] - E\left(\sum_v I_v\right)^2 \\ &= \sum_v \mathbb{E}(I_v^2) + \sum_{\substack{(v,v')\\ v\neq v'}} \mathbb{E}(I_v I_{v'}) - (\rho n + o(n))^2.\end{aligned}$$

We focus on the second summation, which is over distinct vertices. Choose distinct vertices v, v' in our graph $\mathcal{G}_{n,p}$. Consider a BRP with v as a root. With probability $\rho + o(1)$, vertex v is in a small component and in this case $T \leq k_-$. With probability $1 - o(1)$, vertex v' has not yet been discovered in the process. Then consider a separate BRP beginning at $v' \neq v$ on the graph obtained by removing the component containing v. Since we have only removed $O(\log n)$ vertices, this new graph has a number of vertices which is on the order of n. Hence the probability that v' is in a small component is again $\rho + o(1)$. This gives us that $\mathbb{E}(I_v I_{v'}) = \mathbb{P}(I_v = I_{v'} = 1) = \rho^2 + o(1)$ for $v \neq v'$. Therefore

$$\mathrm{Var}(X) \leq n(\rho + o(1)) + n^2(\rho^2 + o(1)) - (\rho^2 n^2 + o(n^2)) = o(n^2),$$

completing the proof. □

We may now prove **Theorem 3.11.**

Proof. By Lemma 3.13, we know that if a component of size at least k_+ exists, then it is unique. Moreover, by Lemma 3.12 we have that the second largest component is of size at most k_-. Recall from Lemma 3.14 that X denotes the number of vertices of $\mathcal{G}_{n,p}$ in components of size at most k_-. Then applying Chebyshev's inequality (Corollary 2.16) followed by Lemma 3.14, we have that for any $\epsilon > 0$,

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geq \epsilon \mathbb{E}(X)\right) \leq \frac{\mathrm{Var}(X)}{\left(\epsilon \mathbb{E}(X)\right)^2} = \frac{o(n^2)}{\epsilon^2 \rho^2 n^2 + o(n^2)} = o(1).$$

Therefore, as $\mathbb{E}(X) = \rho n + o(n)$, we have that $X = \rho n + o(n)$ with probability $1 - o(1)$. Hence w.h.p., the number of vertices in the unique giant component is $n - \rho n + o(n)$, and therefore the fraction of vertices in the giant component is $(1 - o_p(1))\beta n$, completing the proof. □

Our arguments in the proofs of Theorems 3.10 and 3.11 relied heavily upon looking at the expected number of neighbours of a randomly chosen vertex, that is, its expected degree. We found that when the expected degree of a vertex exceeds 1, the giant component emerges. In the next chapter, we consider the emergence of the giant component in random graphs where the degree of each vertex is specified.

Chapter 4

The Giant Component for an Arbitrary Degree Distribution

4.1 Degree Distributions

When it comes to real-world graphs (often called *networks* in the physics and social science literatures), there are many properties which have been observed that the binomial random graph $\mathcal{G}_{n,p}$ simply does not exhibit. Thus in practice, knowing the threshold for the emergence of the giant component is only useful when dealing with very specialised types of networks. We would like to know when a giant component exists in a much more general setting. Homophily is the idea that people have connections with those who are of a similar "type" to them. This has been observed substantially in a variety of friendship and relationship networks ([35,40]), and this property has an effect on diffusive processes that can take place on networks [25,49] (such as the kind of process we will look at in Chapter 5). This property is not captured by $\mathcal{G}_{n,p}$. We say that there is *clustering* in a graph when the graph has small subgraphs with lots of edges. We observe significantly more clustering in real world networks than we observe in the binomial random graph [56]. Indeed $\mathcal{G}_{n,p}$ is quite sparse when $p = o(1)$, in the sense that we would not expect to find edges concentrated around a few vertices, but rather spread out over the whole graph. There are a plethora of other observed properties we could talk about here that are well studied, such as small average path length, small diameter, and low density ([40, 52, 56]), but we have chosen to focus on the *distribution of degrees* of vertices in a graph.

Definition 4.1. Let G be a random graph on $[n]$. The *degree distribution* $\mathcal{D}$ is defined by the probability density function $\{p_k\}_{k\in\mathbb{N}}$ such that $\mathbb{P}(\deg(v) = k) = p_k$ for any vertex $v \in [n]$ chosen uniformly at random.

We will say that a vertex is "chosen at random" when it has been chosen uniformly at random from the set of all vertices. Put simply, the degree distribution tells us the probability that a randomly chosen vertex has a particular degree. The degree distribution certainly plays an important role in real-world networks. For example real-world networks often exhibit the Pareto principle: the idea that for many events, roughly 80% of the effects arise from only 20% of the causes [45]. This translates to finding a few nodes with very high degree, and the rest of the nodes with relatively small degree. Initial attempts attempts to model this precisely made use of the so-called *power law* distribution [1,7,48], however recent evidence indicates that other heavy tailed distributions may be a better fit for the data [12, 14].

The many properties of interest for a random graph are usually *implicitly* defined as a consequence of specifying the procedure by which the graph is constructed. Therefore it is only natural that if we wish to introduce degree distributions into our study of random graphs, we must alter the procedure by which our graphs are constructed. Indeed, we would like to extend our analysis of the giant component in some natural way such that we can apply it to a random graph with *any* degree distribution. We will first look at the degree distribution for $\mathcal{G}_{n,p}$.

4.1.1 Degree Distribution for $\mathcal{G}_{n,p}$

Consider the binomial random graph $\mathcal{G}_{n,p}$. For a randomly chosen vertex v of $\mathcal{G}_{n,p}$, and a fixed $k \in \mathbb{N}$,

$$p_k = \mathbb{P}(\deg(v) = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

So the degree distribution of $\mathcal{G}_{n,p}$ is $\mathcal{D} = \mathrm{Bin}(n-1, p)$. This is why $\mathcal{G}_{n,p}$ is called the *binomial* random graph. If $np \to \lambda$ as $n \to \infty$, then using (2.5) to estimate the binomial coefficient we have that

$$\begin{aligned}
\lim_{n\to\infty} \binom{n-1}{k} p^k (1-p)^{n-1-k} &= \lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \lim_{n\to\infty} \frac{n^k}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{-k} \left(1-\frac{\lambda}{n}\right)^n \\
&= \lim_{n\to\infty} \left(\frac{n^k}{k!} \cdot \frac{\lambda^k}{n^k}\right) \left(1-\frac{\lambda}{n}\right)^{-k} \left(1-\frac{\lambda}{n}\right)^n \\
&= \frac{\lambda^k}{k!} \cdot 1 \cdot e^{-\lambda}. \qquad (4.1)
\end{aligned}$$

This is simply the familiar fact that the limit of a binomially-distributed random variable with $np = \lambda$ is a Poisson distribution with mean λ. For this reason, the random graph $\mathcal{G}_{n,p}$ is sometimes called the *Poisson* random graph. This is particularly common parlance in the physics and social science literatures. If np does not tend to a constant, then under certain conditions we have that the binomial distribution approaches a normal distribution as $n \to \infty$.

Theorem 4.2 (Barbour, Karoński, and Ruciński, 1989)**.** *Let $S = S_n(d)$ be the number of vertices of degree d in $\mathcal{G}_{n,p}$. Let $W \coloneqq (S - \mathbb{E}(S))/\sqrt{\mathrm{Var}(S)}$. Then for $d \geq 1$, we have that $W \xrightarrow{d} N(0,1)$ if and only if*

$$n^{d+1}p^d \to \infty \quad \textit{and} \quad np - \log n - d\log\log n \to -\infty.$$

A slightly weaker but also sufficient condition for normality is that $\mathbb{E}(S) \to \infty$ and either $np \to 0$ or $np \to \infty$. We now present an algorithm for constructing a random graph with an arbitrary degree distribution. The algorithm creates an auxiliary probability space (random configurations) which is studied in its own right in Section 4.3. The main reference for the following section is [22, Chapter 11].

4.2 The Configuration Model

Let $\mathbf{d} = (d_1, d_2, \ldots, d_n)$ be a sequence of positive integers such that $\sum_{i=1}^n d_i = 2m$ is even. Let

$$\mathscr{G}_{\mathbf{d}} = \{\text{simple graphs with vertex set } [n] \text{ s.t. } \deg(i) = d_i,\ i \in [n]\}$$

and let $\mathbb{G}_{\mathbf{d}}$ be chosen randomly from $\mathscr{G}_{\mathbf{d}}$. We assume that $d_i \geq 1$ (for $i = 1, \ldots, n$, since vertices of degree zero are not of interest and unnecessarily complicate matters, and we also assume that

$$\sum_{i=1}^n d_i(d_i - 1) = \Omega(n),$$

which ensures that the graph is not too sparse.

The following algorithm is due to Bollobás [8], though was independently discovered by Wormald [57] at the same time. Let $\mathbf{d}$ be as above. Let W be a set of elements called *points*, such that $|W| = 2m$. Consider a partition $W_1, W_2, \ldots, W_n$ of W such that $|W_i| = d_i$ for $1 \leq i \leq n$. We call $W_1, \ldots, W_n$ *cells*. We fix a total order $<$ on

W, such that $x < y$ if $x \in W_i$ and $y \in W_j$ with $i < j$. For $x \in W$, define $\varphi(x)$ to be the index such that $x \in W_{\varphi(x)}$. Let F be a partition of W into m pairs (*a configuration*). Given F, we define the (multi)graph $\gamma(F)$ as

$$\gamma(F) = \left([n], \{(\varphi(x), \varphi(y)) \colon (x, y) \in F\}\right).$$

The formal definition is a little abstract and so the easiest way to get a feel for what is going on here is via an example. One can think of the above process as assigning to each vertex i, a number d_i of "stubs" or "half edges". The configuration algorithm then matches these stubs up at random as in the picture we have provided below. Note that technically each stub in Figure 4.1 should have its own label, since we need to distinguish between the configuration which we give in Figure 4.2, and any other configuration which gives an isomorphic graph, but matches the stubs up differently. For example, an isomorphic graph could be obtained by swapping the two stubs which link vertex 1 and 5. This is a technicality, and the image below is given more for intuition than rigour.

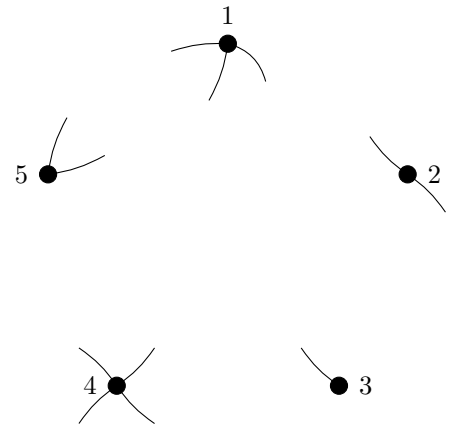


Figure 4.1: Stubs on 5 vertices with $\mathbf{d} = (3, 2, 1, 4, 2)$

Definition 4.3. The *configuration model* is the algorithm (described above) which, for a given degree sequence $\mathbf{d}$, chooses a partition F of W uniformly at random and constructs the (multi)graph $\gamma(F)$. This defines a probability space Ω_n over the set of all possible graphs (configurations) with a given degree sequence $\mathbf{d}$. We will denote a graph generated by the configuration model as $\mathbb{G}^\star_{\mathbf{d}}$ rather than $\gamma(F)$ to remain consistent with the notation for a simple graph $\mathbb{G}_{\mathbf{d}}$ with degree sequence $\mathbf{d}$.

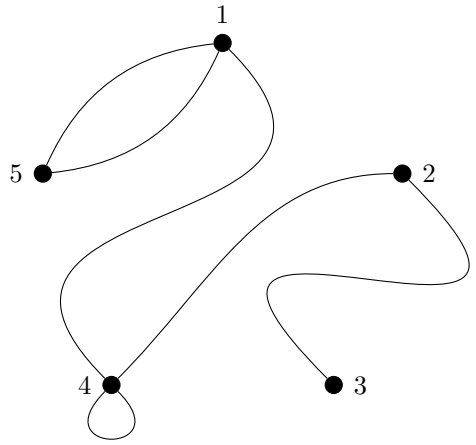


Figure 4.2: Configuration model matches stubs up at random

Let Ω_n denote the set of all configurations defined above for $d_1, \cdots + d_n = 2m$. Since we are working with a uniform probability space over Ω_n, it would be useful to know $|\Omega|$. To find this, consider the sequence

$$(\underbrace{1, 1, \ldots, 1}_{d_1}, \underbrace{2, \ldots, 2}_{d_2}, \ldots, \underbrace{n, \ldots, n}_{d_n}).$$

Take a permutation $(\sigma_1, \sigma_2, \ldots, \sigma_{2m})$ of this sequence, and let F be the partition $F = \{\{\sigma_{2i-1}, \sigma_{2i}\} \colon i = 1, \ldots, m\}$. Note that all possible pairings can be achieved in this way. How many different permutations could have given us this F? Firstly, any permutation which reorders the sequence of pairs $\{\sigma_{2i-1}, \sigma_{2i}\}$ would have given us the same F. There are $m!$ ways to do this. Moreover, any permutation which swaps the order of $\{\sigma_{2i-1}, \sigma_{2i}\}$ would also give the same F. There are 2^m ways to do this (think of a binary switch for each pair). These are all possible rearrangements that would give us F, hence each distinct partition F arises in $m!2^m$ ways. It follows that

$$|\Omega| = \frac{|\{\text{permutations of } 2m \text{ letters}\}|}{|\{\text{partitions which give the same configuration}\}|} = \frac{(2m)!}{m!2^m}.$$

For those comfortable with a bit of group theory, the number of partitions which give the same configuration is the size of the *automorphism group* of the partition.

Next, notice that certain multigraphs are more likely to arise from the configuration procedure than others. An easy way to see this is to consider the case with $n = 2$ and $\boldsymbol{d} = (3, 3)$. There are only two multigraphs with degree sequence $\mathbf{d}$, up to isomorphism, pictured below. The multigraph on the left has $3 \times 3 = 9$ corresponding configurations (consider the edge which joins the two vertices), whereas the graph

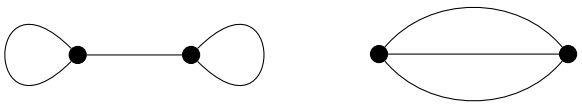


Figure 4.3: Two possible graphs with $\mathbf{d} = (3, 3)$

on the right only has $3! = 6$ corresponding configurations. The probabilities of forming different simple graphs behaves much more nicely. The following relationship holds between a simple graph $G \in \mathscr{G}_{\mathbf{d}}$ and the number of configurations F for which $\gamma(F) = G$.

Lemma 4.4. *If $G \in \mathscr{G}_{\mathbf{d}}$, then*

$$\left|\gamma^{-1}(G)\right| = \prod_{i=1}^{n} d_i! \ .$$

My proof takes a slightly different approach to that offered by Frieze & Karoński [22].

Proof. Suppose that $F \in \Omega_n$ is a configuration such that $\gamma(F) = G$. Consider any point v in the cell W_1. Since G is simple, this point is linked to a point in a different cell W_i. Moreover, no other points in W_1 are linked to any points in W_i. Hence by replacing v with any other point v' in W_1, we have a new configuration which gives the same graph G. There are $|W_1| = d_1$ such choices for v' (including v). Once this is fixed, move on to another point w in W_1. By the same argument as above, we have $d_1 - 1$ choices for replacing this point which will still yield the same G. Continuing in this way, we see that we have $d_1!$ ways to arrange the points in W_1 such that we still have $\gamma(F) = G$. The lemma follows by applying the same argument to $W_2, \ldots, W_n$. □

What is more important than the above lemma is the following corollary.

Corollary 4.5. *If F is chosen uniformly at random from the set of all configurations Ω_n, and $G_1, G_2 \in \mathscr{G}_{\mathbf{d}}$, then*

$$\mathbb{P}\left(\gamma(F) = G_1\right) = \mathbb{P}\left(\gamma(F) = G_2\right).$$

This means that if we run the configuration model and condition on the event "G is simple", then the outcome is uniform over the resulting graphs, that is, over $\mathscr{G}_{\mathbf{d}}$. Clearly this is only a useful exercise if the probability of forming a simple graph is sufficiently large. We state the following theorem without proof, to convince the reader that searching for simple graphs via the configuration model is indeed a plausible strategy when $\mathbf{d}$ is sufficiently sparse.

Theorem 4.6 (Bollobás 1985)**.** *Suppose that* $\Delta = \max\{d_1, \ldots, d_n\} \leq n^{1/7}$*. If* F *is chosen uniformly at random from the set of all configurations* Ω*, then*

$$\mathbb{P}(\gamma(F) \text{ is simple}) = (1 + o(1))\, e^{-\lambda(\lambda+1)},$$

where

$$\lambda = \frac{\sum_{i=1}^{n} d_i(d_i - 1)}{2\sum_{i=1}^{n} d_i}.$$

Moreover, for any (multi)graph property $\mathscr{P}$*,*

$$\mathbb{P}(\mathbb{G}_{\mathbf{d}} \in \mathscr{P}) \leq (1 + o(1))e^{-\lambda(\lambda+1)}\mathbb{P}(\gamma(F) \in \mathscr{P}).$$

Thus if we are interested in properties of $\mathbb{G}_{\mathbf{d}}$, it suffices to look at whether the property holds for the configuration model with degree sequence $\mathbf{d}$. It is worth mentioning that the proof here is quite involved. The modern technique employed for proving Theorem 4.6 is a sort of double counting argument due to Mckay & Wormald [39] known as "switching". This is not the same technique that was used by Bollobás. We also remark that the condition $\Delta \leq n^{1/7}$ can be relaxed quite a bit when looking at the property "$\mathbb{G}_{\mathbf{d}}$ has a giant component" (which we will be studying shortly). This condition has in fact recently been improved to allow $\Delta = o(n)$ so long as $\mathbb{P}(D \geq 3) > 0$ [10].

4.3 The Giant Component in The Configuration Model

It was Molloy & Reed [42, 43] who first derived formulas for the threshold at which the giant component emerges, and its size, in the configuration model. The approach of Molloy & Reed is to "expose" a random configuration one component at a time using a breadth-first-search algorithm reminiscent of a branching process, which they specify in [42]. The extinction behaviour of the process is then analysed using *random walks*, a standard Markov process. Since [42, 43], several improvements

and alternate proofs have been discovered. A heuristic argument was provided by Newman, Stogatz, & Watts [47] using generating functions, but the argument is not fully rigorous. To the best of our knowledge, no fully rigorous proof along these lines has been given. In this section, we describe the argument of Newman et al., and proceed to make it much more rigorous.

We mentioned in Section 4.1 that it important to be able to specify degrees for modelling real-world networks. From a practical perspective, there have been various attempts in the study of observed networks to estimate degree distributions ([5, 12, 58]). The first and most obvious way to do so is to collect data on a particular type of network and use this to estimate the degree distribution directly.. A more sophisticated, though less falsifiable, method is to develop a model of network formation and to look at the properties of the model ([7, 13, 56]). Ideally one would like to check the predictions of their model against some data, though this can be difficult as network data is costly to collect [2].

From a theoretical perspective, there are two approaches to studying the properties of graphs with an arbitrary degree distribution. The first approach is to study the properties of graphs with a given (fixed) degree distribution. This approach is common among mathematical research. The idea is to fix a sequence of degree sequences $\{\mathbf{d}_n\}_{n\in\mathbb{N}}$ which converges to the desired distribution (in a precise sense), and to look at the asymptotic properties of a uniformly random graph with degree sequence $\mathbf{d}_n$. This is $\mathbb{G}_{\mathbf{d}}$. The second approach, which is more common in physics research, is to first draw the degree of each vertex independently from a given probability distribution on the non-negative integers, and use this random sequence $(d_1, \ldots, d_n)$ as your degree sequence (drawing again if the sum of degrees is odd). Then choose a graph uniformly at random from the set of graphs with this degree sequence. Note that there are is no a priori upper bound on the degree of a vertex in a multigraph, though for there to be any positive probability of constructing a simple graph, we must have $d_i \leq n-1$ for all $i = 1, \ldots, n$. Newman et al. state without proof that when looking at properties of these graphs, the two methods are equivalent in the limit as $n \to \infty$. The first approach is sometimes referred to as the *microcanonical ensemble* for random graphs, whereas the second approach is referred to (in the physics literature) as the *canonical ensemble*. We focus on the canonical ensemble in what follows. In the process of making the argument of [47] rigorous, we provide an original proof that the canonical ensemble is a special case of the

microcanonical ensemble. We will revisit the microcanonical ensemble in Chapter 5 where we discuss an application of our work.

There is one last distinction to be made here between simple graphs and the configuration model. Note that Theorem 4.6 demonstrates the relationship between properties of simple graphs with a given degree sequence, and graphs generated by the configuration model with the same degree sequence. In particular, graph properties do not hold with the same probability for the two types of graphs just described. However, if the probability of having some property vanishes for the configuration model (with a given degree sequence), then it also vanishes for simple graphs with that degree sequence. In other words, we can always pull a "negative" result across from the configuration model to $\mathbb{G}_{\mathbf{d}}$, but we cannot necessarily pull a positive result across. Newman et al. [47] have a tendency to blur the line between these two types of graphs, and never explicitly mention the configuration model.

To avoid ambiguity, we state explicitly here that throughout the remainder of this chapter we will be working with the configuration model $\mathbb{G}^{\star}_{\mathbf{d}}$ on a set of vertices $V = [n]$, where the degree sequence $\mathbf{d} = (d_1, \ldots, d_n)$ is constructed by drawing d_i ($i \in [n]$) independently from a distribution $\mathcal{D}_n \in \{\mathcal{D}_n\}_{n\in\mathbb{N}}$ on the non-negative integers, conditioned on an even sum of the degrees. We will call $\mathcal{D}_n$ (for any fixed n) the degree distribution, and write its p.m.f. as $\{p_k\}_{k\in\mathbb{N}}$, suppressing the dependence on n. The p.m.f. determines the generating function for $\mathcal{D}_n$ (see Remark 2.19). We assume that $\mathcal{D}_n$ converges in distribution to some well-defined distribution $\mathcal{D}$, which we will call the *limiting degree distribution*. Since we are only concerned with properties of G as $n \to \infty$, we can work with $\mathcal{D}$ rather than $\mathcal{D}_n$ without losing any generality in our results. Defining $\mathcal{D}_n$ is more of a formality than a necessity, in what follows we will work mainly with $\mathcal{D}$ and we may occasionally blur the distinction between the two. We note that the authors of [47] do not mention $\mathcal{D}_n$ in their paper, but we have defined it here to provide more rigour to their argument.

4.3.1 Generating functions

Let $G = \mathbb{G}^{\star}_{\mathbf{d}}$ with $\mathbf{d}$ as above. Then the probability that a randomly chosen vertex of G has degree k is given by p_k. Newman et al. do not put any explicit restrictions on the limiting distribution $\mathcal{D}$, but their argument assumes that $\mathcal{D}$ has finite expectation which is bounded away from zero. We reserve the expectation operator $\mathbb{E}$ for distributions whose realisations are graphs, and so we will write $\langle k \rangle := \mathbb{E}(\mathcal{D})$ for

the expected degree of a vertex. Recall that the generating function of $\mathcal{D}$, which we will call $G_0(z)$ is defined by

$$G_0(z) = \sum_{k=0}^{\infty} p_k \, z^k. \tag{4.2}$$

This function has a number of properties which were outlined in Section 2.2. In particular we will make use of Theorem 2.22 which describes how generating functions behave when taking a random number of draws of some random variable.

Example 4.7. *For a randomly chosen vertex $v \in V$ and the graph $\mathcal{G}_{n,p}$, one has that*

$$\mathbb{P}(\deg(v) = k) = \binom{n-1}{k} p^k (1-p)^{n-k}, \tag{4.3}$$

and hence the generating function for the degree distribution of a binomial random graph is

$$G_0(z) = \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-k} z^k = (px + 1 - p)^{n-1} \tag{4.4}$$

as seen in Example 2.21. Note that the sum is cut off at $n-1$ because $\mathcal{G}_{n,p}$ is a simple graph and therefore a vertex can have at most $n-1$ neighbours. This does not have to be the case for the configuration model, unless of course we want to be able to infer something about simple graphs (which we usually do).

Another distribution that will prove useful to us is the degree distribution of a vertex found from choosing an *edge* at random and following it to one of its ends. The probability that a randomly chosen vertex has degree k is p_k, and there are k edges along which this vertex could be reached by following a randomly chosen edge. After normalising, it follows that if v is a vertex found by following a randomly chosen edge, then

$$\mathbb{P}(\deg(v) = k) = \frac{k p_k}{\sum_j j p_j}. \tag{4.5}$$

Hence the *edge-following* generating function is

$$\sum_{k=0}^{\infty} \left(\frac{k p_k}{\sum_j p_j} \right) z^k = \frac{G_0'(z)}{G_0'(1)} \, z. \tag{4.6}$$

Now consider choosing a vertex u at random and following one of its incident edges to a previously unexplored neighbour w. The degree distribution of this neighbour

is exactly the same as that of choosing an edge at random and following it to one of its ends. In this setting we are interested in the number of neighbours that a vertex $w$is connected to, other its "predecessor" u. This is called the *forward degree* of the vertex, and the corresponding distribution is called the *forward degree distribution*, denoted $\mathcal{D}'_n$. One can imagine why this is a quantity of interest: if we wish to use a branching process to find the asymptotic size of a component then at each step of the process we are only interested in how many *more* neighbours a particular vertex has, other than the ones we have already visited. If the forward degree of a vertex is k, then the degree of the vertex is $k+1$ (by counting the predecessor as well). Thus the forward degree distribution is defined by

$$q_k = \frac{(k+1)p_{k+1}}{\sum_{j=0}^{\infty}(j+1)p_{j+1}} = \frac{(k+1)p_{k+1}}{\sum_{j=0}^{\infty} jp_j}, \tag{4.7}$$

which has generating function

$$\sum_{k=0}^{\infty}\left(\frac{(k+1)p_{k+1}}{\sum_j p_j}\right)z^k = \frac{\sum_{k=0}^{\infty} kp_k z^k}{\sum_j jp_j}\cdot\frac{1}{z} = \frac{G'_0(z)}{G'_0(1)}. \tag{4.8}$$

Here we assume that assuming that $z \neq 0$ in order to divide throughout by it, though taking the limit as $z \to 0$ will agree with the above formula. Moreover, noting that $G'_0(1) = \langle k\rangle$ is the first moment of the distribution (that is, the average degree of a vertex), we can write the generating function for the forward degree distribution as

$$G_1(z) \coloneqq \frac{1}{\langle k\rangle}G'_0(z). \tag{4.9}$$

We see that $G_1(z)$ tells us the generating function for the number of "first (forward) neighbours" of a vertex v. What about the number of *second* neighbours, that is, neighbours of neighbours? What about *third* neighbours, and so on? This is a question we will return to after our more rigorous treatment of this section.

We now turn to the question of the threshold for the emergence of the giant component.

4.3.2 Component sizes

Let us introduce now a new distribution of interest: the distribution over the sizes of connected components in the graph $\mathbb{G}^{\star}_{\mathbf{d}}$ as $n \to \infty$ *not including vertices in components of infinite size* (if there are any). We will make this precise in a moment, but before proceeding further, we remark here that Newman et al. are less careful

with their description than we have been here. They exclude vertices in "the giant component" from this distribution, but excluding vertices in components of infinite size is what is actually needed in order to apply their methodology. The distinction is important, since the method does not prove that these vertices are in a *single common component*. Hence we more carefully state "vertices in components of infinite size". Indeed it will turn out that these vertices will in fact lie in a unique giant component; we will return to this question at the end of Section 4.3.3.

Let $G = \mathbb{G}^{\star}_{\mathbf{d}}$ and let $\mathbf{d}$ be a degree sequence drawn from $\mathcal{D}_n$. For $n \in \mathbb{N}$, let $c_{n,k}$ be the probability that a randomly chosen vertex $v \in \mathbb{G}^{\star}_{\mathbf{d}}$ is in a component of size $k \in \mathbb{N}$, and let $c_k := \lim_{n\to\infty} c_{n,k}$. So $\{c_k\}_{k\in\mathbb{N}}$ the distribution over the sizes of connected components in the graph $\mathbb{G}^{\star}_{\mathbf{d}}$ as $n \to \infty$. Then define

$$H_0^{(n)}(z) = \sum_{k=0}^{n} c_{n,k} z^k$$

to be the generating function for the distribution of sizes of components in $\mathbb{G}_{\mathbf{d}}$. The limiting generating function for the distribution of component sizes is given by

$$H_0(z) = \lim_{n\to\infty} H_0^{(n)}(z) = \lim_{n\to\infty} \sum_{k=0}^{n} c_{n,k} z^k = \sum_{k=0}^{\infty} c_k z^k.$$

The function H_0 is the main object of interest in this section. By excluding vertices in "components of infinite size" from $\{c_k\}_{k\in\mathbb{N}}$, we really mean that $H_0^{(n)}(1) = 1 - o(1)$. That is, that $\lim_{n\to\infty} H_0^{(n)} = 1$, or equivalently

$$H_0(1) = \sum_{k=0}^{\infty} c_k = 1. \tag{4.10}$$

We require that (4.10) holds because we want to ensure we are in the "subcritical" regime, where all branching processes go extinct with probability 1. The necessity of this condition will become clear shortly, as we will assume that $\mathbb{G}^{\star}_{\mathbf{d}}$ is tree-like and is therefore well approximated by a branching process.

We introduce one other generating functions before we proceed. Consider again choosing a random edge and following it to a vertex which is incident with it. Let $H_1^{(n)}(z)$ and $H_1(z)$ be the generating function for the distribution of sizes of components *reached in this way*, defined analogously to $H_0^{(n)}$ and H_0. The way that we have defined H_0 and H_1 here differs slightly from Newman et al. [47]. This is

because they do not consider $\mathcal{D}_n$ in their paper, but rather work immediately with the limiting distribution $\mathcal{D}$. This will not affect our analysis of component sizes.

Assume for now that G is a tree, or at least "locally" a tree (we make this precise in Lemma 4.12 in Section 4.3.4). Then by following a randomly chosen edge, we reach a single vertex v_0, plus any number of other treelike "clusters" with the same size distribution, joined to this vertex by single edges. Newman et al. represent this idea pictorially as below. Let q_k be (as before) the probability that the initial

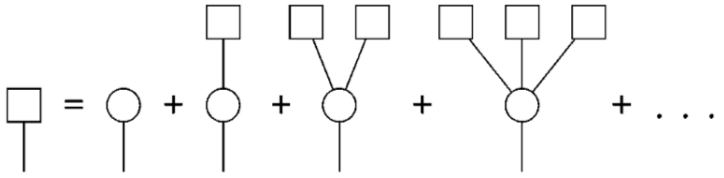


Figure 4.4: Figure taken from [47]. Pictorial representation of following a randomly chosen edge to vertices and tree-like clusters.

vertex we find has forward degree k. Then a forward degree of k for our initial vertex v_0 will result in k random draws from the distribution generated by $H_1(z)$. It follows by applying Theorem 2.22 that $H_1(z)$ must satisfy

$$H_1(z) = z \sum_{k=0}^{\infty} q_k \left(H_1(z)\right)^k = z\, G_1(H_1(z)). \tag{4.11}$$

This has been called a *self-consistency condition* and should be reminiscent of finding the extinction probability of a branching process. Rather than using Theorem 2.22 one can alternatively look at Figure 4.4 to see almost immediately that we must have

$$H_1(z) = z\, q_0 + z\, q_1 H_1(z) + z\, q_2 [H_1(z)]^2 + z\, q_3 [H_1(z)]^3 + \ldots$$

which yields the same result.

If we begin a BRP at a randomly chosen vertex, then each edge incident with the vertex would give us one treelike "cluster", and so by the same reasoning as above,

$$H_0(z) = z\, G_0(H_1(z)). \tag{4.12}$$

Remember that $H_0(z)$ is the function we are most interested in, as it tells us the distribution of component sizes. If we have the generating function $G_0(z)$ for the degree distribution, then (at least in theory) we can calculate $G_1(z)$ and solve (4.11) to get $H_1(z)$, finally substituting this into (4.12) to obtain $H_0(z)$. Note that this entire process depends on knowledge of the degree distribution and therefore highlights the need to be being able to estimate degree distributions. The probability that a vertex chosen at random belongs to a component of size s is then the coefficient of z^s in H_0; that is, the s-th derivative of H_0, evaluated at 0. Calculating this value is of course a computationally laborious and sometimes impossible exercise for large s, though we can approximate these derivatives by numerical integration of the Cauchy formula, giving us the probability distribution P_s of component sizes:

$$P_s = \frac{1}{s!}\frac{d^s H_0}{dz^s}\bigg|_{z=0} = \frac{1}{2\pi i}\oint \frac{H_0(z)}{z^{s+1}}\,dz. \tag{4.13}$$

This integral will contain no poles of the generating function for $|z| \leq 1$ (and possibly larger values) and thus can always be taken over the contour $|z| = 1$. If it is possible to take larger contours then they will yield better numerical approximations, but it is not always possible to take them.

Example 4.8 (Original)**.** *For a positive integer k we write $(k)!! = k(k-2)(k-4)\cdots j$, where $j = 1$ if k is odd, and $j = 2$ if k is even. Suppose $G(z) = \frac{2}{\pi}\arcsin z = \frac{2}{\pi}\sum_{k=0}^{\infty}\frac{(2k-1)!!}{(2k)!!}\frac{z^{2k+1}}{2k+1}$. Then the largest contour over which $G(z)$ is integrable is the unit circle (the limit as $z \to 1$ from the right does not exist).*

4.3.3 The threshold for the emergence of the giant component

We have already established that it is difficult to find a closed form expression for (4.11) and (4.12). However, it is possible to find a closed form solution for the *average* size of the component in which a randomly chosen vertex is contained. If there are no vertices in components of infinite size, then we can simply use H_0 and calculate the average as usual:

$$\langle s\rangle = H_0'(1) = G_0(H_1(1)) + G_0'(H_1(1))H_1'(1) = 1 + G_0'(1)H_1'(1). \tag{4.14}$$

The final inequality comes from the fact that $G_0(H_1(1)) = H_0(1)$ (by (4.12)), and $H_0(1) = 1$ because the coefficients of z^k in H_0 form a probability distribution. We

have from (4.11) that

$$H_1'(1) = 1 + G_1'(1)H_1'(1) = \frac{1}{1 - G_1'(1)}, \tag{4.15}$$

and hence

$$\langle s \rangle = 1 + \frac{G_0'(1)}{1 - G_1'(1)}. \tag{4.16}$$

We have by definition of G_1 that

$$\left[\frac{d}{dz}G_1(z)\right]_{z=1} = \frac{1}{\langle k \rangle}G_0'\prime(1) = \frac{G_0'\prime(1)}{G_0'(1)}. \tag{4.17}$$

Therefore we can write (4.16) as

$$\langle s \rangle = 1 + \frac{(G_0'(1))^2}{G_0'(1) - G_0''(1)}. \tag{4.18}$$

We can see that this equation for $\langle s \rangle$ diverges when $G_0'(1) = G_0''(1)$. But this identity holds precisely when the expected number of first neighbours of a vertex is equal to the expected number of second neighbours! That is, we find exactly what we might have expected in considering a branching process; if the expected number of offspring exceeds 1, the process will continue forever and the component size will be infinite. This may not be immediately clear but we will discuss it further in Section 4.3.5 (see (4.30)). Since

$$G_0'(1) = \sum_{k=0}^{\infty} kp_k \quad \text{and} \quad G_0''(1) = \sum_{k=0}^{\infty} k(k-1)p_k,$$

the expected size of a component diverges to infinity if

$$\sum_{k=0}^{\infty} kp_k = \sum_{k=0}^{\infty} k(k-1)p_k \tag{4.19}$$

that is, when,

$$\sum_{k=0}^{\infty} k(k-2)p_k = 0. \tag{4.20}$$

Note that from (4.18), the average component size is finite if and only if $G_0'(1) > G_0''(1)$. It follows that as $n \to \infty$, vertices in components of infinite size exist if and only if $G_0'(1) > G_0''(1)$, or equivalently (by (4.20)), when $\sum_{k=0}^{\infty} k(k-2)p_k > 0$.

Remarkably, this is exactly the threshold for the emergence of the giant component that Molloy & Reed (1995) found using their rigorous argument.

Recall that Newman et al. excluded "a giant component" from the distribution over component sizes in their analysis, whereas we excluded "vertices in components of infinite size". I provide an original analysis of these vertices in components of infinite size. If (4.18) diverges, then this means that we expect a randomly chosen vertex to be in a component of infinite size. Let X be a random variable counting the number of vertices in components of infinite size. Then $X = \sum_{v \in V} I_v$ where I_v is an indicator variable for the event "v is in a component of infinite size". Hence

$$\mathbb{E}(X) = \sum_{v \in V} \mathbb{E}(I_v) = n\,\mathbb{P}(v \text{ is in a component of infinite size}). \tag{4.21}$$

Intuitively, if the expected component size is infinite, then we expect that the extinction probability ρ of a BRP beginning at a randomly chosen vertex is strictly less than 1 (Corollary 3.4). It follows that for a randomly chosen vertex v, the probability of survival of the BRP is $1 - \rho$ for some $\rho \in (0, 1)$. Hence

$$\mathbb{P}(v \text{ is in a component of infinite size}) = 1 - \rho > 0$$

and therefore $\mathbb{E}(X)$ is of linear order. So indeed, we must have at least one giant component.

When it exists, the giant component is unique, however the argument required to prove this is quite involved. We mentioned at the beginning of the chapter that Molloy & Reed use a variation of exploring a component in the configuration model based loosely on branching processes. Their proof of uniqueness of the giant component can be found in [42, Lemma 11]. The approach of Bollobás & Riordan [10] applies a colouring and sprinkling technique based on the two-round exposure method described in Section 2.4.4. Other approaches have been taken to prove that the giant component is unique when it exists [21, 31, 33], though each takes several pages to explain rigorously. Thus we do not prove here that the giant component is unique: we simply highlight that this is true and refer the reader to the above sources should they wish to see the argument for themselves. Whilst the sprinkling method of Bollobás & Riordan is beyond the scope of this thesis, these authors provide an excellent rigorous treatment of the threshold for the emergence

of vertices in components of linear order. We present part of their treatment below, as it allows us to deal with some of the unjustified assumptions of the above argument.

4.3.4 Formal analysis of BRP's with the configuration model

The following analysis is based on Bollobás & Riordan [10]. At the beginning of Section 4.3, we introduced the *canonical* and *microcanonical* ensembles. The authors provide their results within the microcanonical ensemble, but we have reframed the results in terms of the canonical ensemble so that we may be consistent with our analysis of [47]. We will need to make use of some of the features of the canonical ensemble to provide an analogous but not identical statement to [10, Lemma 4]. First, we provide a lemma of our own establishing some asymptotic properties of the canonical ensemble.

Lemma 4.9 (Original)**.** *Let $G = \mathbb{G}^\star_{\mathbf{d}}$. Let $n_k(\mathbf{d})$ be the (random) number of vertices of degree k in $\mathbf{d}$, and $m(\mathbf{d})$ the (random) number of edges in G. Then w.h.p.*

$$\frac{n_k(\mathbf{d})}{n} = p_k + o(1), \tag{4.22}$$

and

$$\frac{m(\mathbf{d})}{n} = \frac{\langle k \rangle}{2} + o(1). \tag{4.23}$$

Proof. Both results follow almost immediately from Hoeffding's inequality (Theorem 2.40). We first prove (4.22). Note that since each vertex has degree k independently with probability p_k, the number of vertices of degree k is binomially distributed with success probability p_k. Let I_v be an indicator variable for the event that vertex v has degree k. Then $n_k(\mathbf{d}) = \sum_{v \in V} I_v$ and so the average $\overline{n_k(\mathbf{d})}$ is given by $\frac{1}{n}\sum_{v \in V} I_v$. Therefore by linearity of expectation we have that $\mathbb{E}(\overline{n_k(\mathbf{d})}) = p_k$. Moreover, since each I_v is bounded on $[0, 1]$, we have by Hoeffding's inequality (Theorem 2.40) that for any $t > 0$,

$$\mathbb{P}\left(\left|\frac{n_k(\mathbf{d})}{n} - p_k\right| \geq t\right) \leq 2\exp\left(-2nt^2\right) = o(1).$$

proving (4.22).

The proof of (4.23) follows from (4.22). Notice that $m(\mathbf{d}) = \frac{1}{2}\sum_{k=0}^{\infty} k n_k(\mathbf{d})$. Hence the average, $\overline{m(\mathbf{d})}$, is given by

$$\overline{m(\mathbf{d})} = \frac{1}{n}\left(\frac{1}{2}\sum_{k=0}^{\infty} k n_k(\mathbf{d})\right) = \frac{1}{2}\sum_{k=0}^{\infty} k\frac{n_k(\mathbf{d})}{n}.$$

By (4.22), we have that

$$\overline{m(\mathbf{d})} = \frac{1}{2}\sum_{k=0}^{\infty} k p_k + o(1) = \frac{\langle k\rangle}{2} + o(1),$$

proving (4.23). □

Lemma 4.9 can now be applied to give an asymptotic expression for the ratio of the number of vertices of degree k to the total degree.

Corollary 4.10 (Original)**.** *Let $n_k(\mathbf{d})$ and $m(\mathbf{d})$ be as above. Then w.h.p.*

$$\frac{n_k(\mathbf{d})}{2m(\mathbf{d})} = \frac{p_k}{\langle k\rangle} + o(1). \tag{4.24}$$

In words, (4.22) says that $\mathcal{D}$ captures the asymptotic proportion of vertices of a certain degree, and (4.23) says that the number of edges in the graph is related to $\mathcal{D}$ in the natural way.

We now establish that searching for the $t-$th neighbours of a vertex is well approximated by a branching process. Bollobás & Riordan have an alternative, but equivalent, characterisation of branching processes which will prove useful, so we give it here. We use notation from Section 3.1.4.

Definition 4.11. Let $\{G_i\}_{i\in\mathbb{N}}$ be a (not necessarily uniform) BRP on $\{X_i\}_{i\in\mathbb{Z}^+}$. A *random rooted tree* $\mathcal{T}$ is the (random) graph determined by this branching process in the natural way, with G_0 distinguished from the other vertices as the "root". That is, by assigning a vertex to each offspring born in the process, and assigning an edge between every parent and their offspring.

Let $\mathcal{T} = \mathcal{T}_{\mathcal{D}}$ be a random rooted tree on $X_1, X_2, \ldots$ with $X_i \sim \mathcal{D}'$, the forward degree distribution (independently for all i). Our goal is to show that this is a

"good" approximation of the BRP beginning at a random vertex of a graph $\mathbb{G}^\star_{\mathbf{d}}$ generated by the configuration model. To do this, we first make the following adjustment. Suppose that X_1 is distributed according to $\mathcal{D}$ rather than $\mathcal{D}'$, so that the tree mirrors the process of choosing a random vertex in a graph with degree sequence drawn from $\mathcal{D}$, and following the edges with which it is incident. The reader may be wondering why we did not have to make this adjustment when proving our results for $\mathcal{G}_{n,p}$ in Chapter 3. The intuitive answer is that since the limiting degree distribution of $\mathcal{G}_{n,p}$ is Poisson (see (4.1)), one can easily verify that $G_0(z) = G_1(z)$ (now going back to Examples 3.5 and 3.6). So, in fact, $\mathcal{G}_{n,p}$ provides a simple case where the forward degree distribution is equal to the degree distribution in the limit.

Next we present the lemma establishing that the configuration model produces a locally tree-like graph. Given a graph G, let $\Gamma_{\leq t}(v) = \Gamma^G_{\leq t}(v)$ denote the subgraph of G induced by the vertices within distance t of v; that is, up to the "t-th neighbours" of v. Similarly, let $\mathcal{T}_{\mathcal{D}}|_t$ be the subtree of $\mathcal{T}_{\mathcal{D}}$ induced by the vertices within distance t of the root (that is, the first t generations of the process).

Lemma 4.12 (Bollobás & Riordan, 2015)**.** *Let v be a vertex of $G = \mathbb{G}^\star_{\mathbf{d}}$ chosen uniformly at random. Then we may couple the random graphs $\Gamma_{\leq t}(v)$ and $T_{\mathcal{D}}|_t$ so that they are isomorphic as rooted graphs with probability $1 - o(1)$ as $n \to \infty$.*

Before we offer a formal proof, we provide an original argument to give some intuition as to why this is the case. Really what we want to show is that a BRP beginning at a randomly chosen vertex of G finds the right number of new vertices in each generation. We will show that this is indeed the case for the first generation and then provide the proof of the lemma. Consider the k different points in the cell corresponding to v. If these points loop back to v with too high a probability, then we may not get a new vertex in the first step as we explore G. We show this happens with probability $o(1)$. Consider the first stub of v. If v has degree k, then the probability that the first stub was linked to another stub of v by the configuration procedure (hence forming a loop) is

$$\mathbb{P}(\text{first stub of } v \text{ forms a loop} \mid \deg(v) = k) = \frac{k-1}{2m-1}.$$

Hence the probability that the first stub of v forms a loop is

$$\sum_{k=1}^{n}\left(\frac{k-1}{2m-1}\right)p_k = \frac{1}{2m-1}\left(\sum_{k=1}^{n}kp_k - \sum_{k=1}^{n}p_k\right)$$
$$= \frac{\langle k\rangle - 1 + p_0}{2m-1}.$$

Now, by (4.23) in Lemma 4.9, we have that

$$\frac{\langle k\rangle - 1 + p_0}{2m-1} = \frac{\langle k\rangle - 1 + p_0}{n\langle k\rangle + o(n) - 1} \leq \frac{C}{n} = o(1),$$

for a large enough positive constant C. Since the first point of the cell corresponding to v gives us the largest number of choices for a potential loop, the probability that any of the other points form a loop is also $o(1)$. This establishes that at every step when we look at the neighbours of v, we expect them to be new neighbours with probability $1 - o(1)$. We now present the proof of Lemma 4.12 which uses the *coupling* technique defined in the preliminaries (see Definitions 2.28 and 2.31).

Proof of Lemma 4.12. By definition, a randomly chosen vertex has degree distribution given by $\mathcal{D}_n$. Since $\mathcal{D}_n \xrightarrow{d} \mathcal{D}$, we can find a vertex $v \in V(G)$ which has the same degree distribution as the root of $\mathcal{T}_{\mathcal{D}}$ with probability $1 - o(1)$. From here, the idea is to reveal the vertices in $\Gamma_{\leq t}(v)$ and $\mathcal{T}_{\mathcal{D}}|_t$ simultaneously, one vertex at a time, and show that the coupling fails with probability $o(1)$ at any given step. Let $R_{k,j}$ denote the number of vertices of degree k which have been *revealed* at the j-th step of looking at unpaired stubs; meaning that we have found these vertices in exploring $\Gamma_{\leq t}(v)$. Moreover, as in Lemma 4.9, let $n_k(\mathbf{d})$ be the number of vertices of degree k in $\mathbf{d}$. At the j-th time we reveal the partner of a single unpaired point in a cell, the probability that this is a "new" vertex of degree k, is precisely

$$\frac{k\left(n_k(\mathbf{d}) - R_{k,j}\right)}{2m+1-2j}. \tag{4.25}$$

Note that for any finite j, we have that $R_{k,j} \leq j = O(1)$. Then making use of both (4.22) and (4.23) in Lemma 4.9, we see that

$$\frac{k\left(n_k(\mathbf{d}) - R_{k,j}\right)}{2m+1-2j} = \frac{knp_k - kR_{k,j}}{n\langle k\rangle + 1 - 2j} + o(1)$$

$$= \frac{kp_k - \frac{kR_{k,j}}{n}}{\langle k \rangle + \frac{1-2j}{n}} + o(1) \tag{4.26}$$

$$= \frac{kp_k}{\langle k \rangle} + o(1).$$

That is, that at step j, the expected number of new vertices of degree k is simply $q_{k-1}+o(1)$. This is asymptotically equal to the forward degree distribution $\mathcal{D}'$ which is also the offspring distribution for $\mathcal{T}_{\mathcal{D}}$ (except for the root which we already dealt with separately). It follows that the coupling succeeds at step j with probability $1-o(1)$. To complete the proof, note that for any $\epsilon > 0$, there is a constant M_ϵ such that with probability at least $1-\epsilon$, the size of the finite tree $\mathcal{T}_{\mathcal{D}}|_t$ is at most M_ϵ. Thus for any $\epsilon > 0$, the probability that the coupling fails is bounded above by the probability that the coupling fails in the first M_ϵ steps, and that $|\mathcal{T}_{\mathcal{D}}|_t| \leq M_\epsilon$. That is, for n large enough, the probability that the coupling fails for the neighbours within distance t of the root is bounded above by $\epsilon + M_\epsilon o(1) = o(1)$, completing the proof. □

The above lemma gives us the following corollary.

Corollary 4.13. *Let v be a vertex of $G = \mathbb{G}^\star_{\mathbf{d}}$ chosen uniformly at random. If $t \geq 1$ is a constant, then w.h.p. the neighbourhood $\Gamma_{\leq t}(v)$ of v in G is a tree.*

What Corollary 4.13 tells us is that when exploring the second generation of a BRP beginning at a vertex v in $\mathbb{G}^\star_{\mathbf{d}}$, we are most likely to find vertices that have not been visited yet, and moreover that this pattern continues for any neighbourhood of finite size. This is precisely why Equations (4.11) and (4.12) hold in the Newman et al. [47] heuristic. In order to write down Equations (4.11) and (4.12) we assumed that G had a locally tree-like structure, which is what we have proved to be the case here.

We continue with a few corollaries which establish that the extinction probability of a branching process captures the asymptotic number of vertices contained in large components. Consider a rooted graph G, and the rooted graph property $\mathscr{P}_k$ that "the component of the root contains exactly k vertices" ($k = 1, 2, \dots$). Remember that we associate a property with the set of rooted graphs (up to isomorphism) which have that property. For a rooted graph property $\mathscr{P}$ we write $(G, v) \in \mathscr{P}$

to mean that the graph G rooted at v has property $\mathscr{P}$. The following is then an immediate consequence of Lemma 4.12.

Corollary 4.14. *Let v be a vertex of $G = \mathbb{G}^\star_{\mathbf{d}}$ chosen uniformly at random. Then*

$$\mathbb{P}\left((G, v) \in \mathscr{P}_k\right) \xrightarrow{d} \mathbb{P}(\mathcal{T}_{\mathcal{D}} \in \mathscr{P}_k)$$

for all $k = 1, 2, \ldots$ as $n \to \infty$. Equivalently, writing $N_k = N_k(G)$ for the (random) number of vertices in G which are in components of size k, we have

$$\mathbb{E}\left(\frac{N_k}{n}\right) = \mathbb{P}(\mathcal{T}_{\mathcal{D}} \in \mathscr{P}_k) + o(1).$$

Proof. The probability that $(G, v) \in \mathscr{P}_k$ depends only on the rooted graph $\Gamma_{\leq t}(v)$ for some sufficiently large t. Since this can be coupled isomorphically with $\mathcal{T}_D|_t$ the first statement follows immediately. The equivalence of the two statements comes from applying linearity of expectation to $N_k = \sum_{v \in V} I_v$ where I_v is the indicator variable for the event "v is in a component of size k". □

In fact, the above statement will hold more generally for rooted graph property which depends only on a *finite neighbourhood* of a randomly chosen vertex. So both Corollary 4.13 and Corollary 4.14 are special cases of a more general statement. What Corollary 4.14 means is that the average number of vertices in components of size k is given by the probability that the random rooted tree $\mathcal{T}_{\mathcal{D}}$ goes extinct after k steps. Let $\rho_k = \mathbb{P}(\mathcal{T}_{\mathcal{D}} \in \mathscr{P}_k)$ be the probability that $\mathcal{T}_{\mathcal{D}}$ goes extinct after producing exactly k offspring (including the initial vertex) and let $\rho(\mathcal{D}) = \lim_{k \to \infty} \rho_k$ be the probability that $\mathcal{T}_{\mathcal{D}}$ never goes extinct. It will be useful to establish another consequence of Corollary 4.14 before we present our final result for this subsection.

Corollary 4.15 (Original)**.** *Let v be a vertex of $G = \mathbb{G}^\star_{\mathbf{d}}$ chosen uniformly at random. Then for any $\epsilon > 0$, there exists $\delta > 0$ (independent of n) such that for n sufficiently large,*

$$\mathbb{P}\left(|N_k(G) - \rho_k n| \geq \epsilon n\right) \leq e^{-\delta n}.$$

Proof. Fix $\epsilon > 0$. It follows from Corollary 4.14 that there exists a sequence $\{\epsilon_n\}_{n\in\mathbb{N}}$ such that $\epsilon_n \to 0$, and

$$\rho_k = \mathbb{E}\left(\frac{N_k}{n}\right) - \epsilon_n.$$

Take n large enough such that $|\epsilon_n| < \epsilon$, then we have that $\epsilon - |\epsilon_n| > 0$. Hence

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{N_k}{n} - \rho_k\right| \geq \epsilon\right) &= \mathbb{P}\left(\left|\frac{N_k)}{n} - \mathbb{E}\left(\frac{N_k}{n}\right) + \epsilon_n\right| \geq \epsilon\right) \\
&\leq \mathbb{P}\left(\left|\frac{N_k}{n} - \mathbb{E}\left(\frac{N_k}{n}\right)\right| + |\epsilon_n| \geq \epsilon\right) \\
&= \mathbb{P}\left(\left|\frac{N_k}{n} - \mathbb{E}\left(\frac{N_k}{n}\right)\right| \geq \epsilon - |\epsilon_n|\right).
\end{aligned}$$

Here, we have used the triangle inequality to break up the absolute value and move the ϵ_n to the other side of the inequality. Then since $\epsilon - |\epsilon_n| > 0$, we have that by Hoeffding's inequality (Theorem 2.40)

$$\mathbb{P}\left(\left|\frac{N_k}{n} - \rho_k\right| \geq \epsilon\right) \leq 2\exp\left(-2n(\epsilon - |\epsilon_n|)^2\right).$$

Let $\delta' = 2(\epsilon - |\epsilon_n|)^2$. Take n large enough such that $\delta' n \geq 2\log 2$, this is possible because $\epsilon_n \to 0$ as $n \to \infty$. Then letting $\delta = \delta'/2$, we have that

$$2\log 2 \leq \delta' n = 2(\delta' - \delta)n. \tag{4.27}$$

Dividing both sides by 2 and exponentiating, we have

$$2 \leq e^{(\delta-\delta')n} \quad \Rightarrow \quad 2e^{-\delta' n} \leq e^{-\delta n},$$

proving the result. □

We have obtained a concentration result for $N_k(G)$ in the canonical ensemble. It is worth noting that Bollobás & Riordan have to go through significantly more work to get an exponential bound (as in our Corollary 4.15) for the microcanonical ensemble (see [10, Theorem 2]). To our knowledge, we are the first to look at this statement specifically in the canonical ensemble and prove that an exponential bound can be attained in this way. The reason that this is important is because one can use this bound to prove that the same statement holds for simple graphs generated by the configuration model. Theorem 4.6 may give the reader a hint as to why this is, though we do not go into any more detail of the proof here.

Now that we have a better understanding of why the generating function heuristic argument of [47] works, we present a few more results that can be derived using it.

4.3.5 Average number of t-th neighbours

We have now seen that for any constant neighbourhood distance $\leq t$ from an arbitrary vertex, w.h.p. $\Gamma_{\leq t}(v)$ is a tree. We can return then to the question of the number of *second*, third, fourth etc. neighbours of a vertex. Beginning at a randomly chosen vertex v, it follows from Lemma 4.12 that the number of "new" (previously unseen) neighbours of v is (asymptotically) distributed according to $\mathcal{D}$, and the number of new "neighbours of a neighbour" is (asymptotically) distributed according to $\mathcal{D}'$. This means that once we choose a random vertex, we draw an integer $k \sim \mathcal{D}$, and then draw k observations of $\mathcal{D}'$ for the number of second neighbours. Hence by Theorem 2.22, the generating function for the probability distribution over the number of second neighbours of a vertex is

$$G_0(G_1(z)) = \sum_{k=0}^{\infty} p_k \left(G_1(z)\right)^k. \tag{4.28}$$

It follows by similar reasoning that the generating function for the number of third-nearest neighbours will be $G_0(G_1(G_1(z)))$, and in general, the number of t-th neighbours, that is, the neighbours which are exactly distance t away from v, will be generated by

$$G_0(G_1^{(t-1)}(z)) = G_0(\underbrace{G_1(\cdots(G_1}_{t-1}(z))). \tag{4.29}$$

Of course, we can use this formula to compute the moments of these distributions. For example, the average number of second neighbours is given by

$$\left[\frac{d}{dz} G_0(G_1(z))\right]_{z=1} = G_0'(G_1(1))\, G_1'(1) = G_0''(1). \tag{4.30}$$

It is interesting to note that the average number of first neighbours is $G_0'(1)$, and the average number of second neighbours is $G_0''(1)$. This is in fact the extent of this pattern; the average number of third neighbours turns out to be $\left(G_1'(1)\right)^2 G_0'(1)$.

Example 4.16. *We have already seen in Example 2.21 that the p.g.f. for the degree distribution of $\mathcal{G}_{n,p}$ is $G_0(z) = (pz + 1 - p)^{n-1}$. From this we have that $G_0'(1) = p(n-1)$ is the average number of first neighbours, and $p^2(n-2)(n-1)$ the average number of second neighbours.*

4.3.6 Generating Functions for the Giant Compononent

It may come as a surprise that the formalism we have introduced for analysing the size of components via generating functions still works when we have vertices in components of infinite size. In this case, $H_0(z)$ generates the size of components *excluding* components of infinite size. This implies of course that $H_0(1) \neq 1$, since by existence of components of infinite size, there is some constant fraction of vertices not represented by H_0 (see (4.21)). One can prove that these vertices will be in a single common component (see for example [10, Theorem 2]), so for the remainder of this chapter, we will refer to these vertices as being in *the giant component*.

We can remedy the fact that $H_0(1) \neq 1$ as follows. Let β be the fraction of vertices contained in the giant component, then $H_0(1) = 1 - \beta$. This allows us to calculate the size of the giant component from (4.11) and (4.12). Let ρ be the extinction probability of a branching process with offspring distribution $\mathcal{D}'$. We know then from our analysis in Section 3.1 that $\rho := H_1(1)$ is the smallest positive real solution of

$$\rho = G_1(\rho), \tag{4.31}$$

or more explicitly,

$$\rho = \sum_{k=0}^{\infty} \frac{k p_k}{\langle k \rangle} \rho^{k-1}. \tag{4.32}$$

Therefore, using (4.12), we have that the size of the giant component is given by

$$\beta = 1 - H_0(1) = 1 - G_0(H_1(1)) = 1 - G_0(\rho). \tag{4.33}$$

What can we say about the average size of components that are not in the giant component? From (4.12) we have that

$$H_0'(1) = G_0(H_1(1)) + G_0'(H_1(1))\, H_1'(1), \tag{4.34}$$

and hence by (4.11), we conclude that the average size of a component (excluding the giant component) is given by

$$\begin{aligned} \langle s \rangle = \frac{H_0'(1)}{H_0(1)} &= \frac{1}{H_0(1)} \left[G_0(H_1(1)) + \frac{G_0'(H_1(1))\, G_1(H_1(1))}{1 - G_1'(H_1(1))} \right] \\ &= \frac{G_0(\rho)}{1-\beta} + \frac{G_0'(\rho) G_1(\rho)}{(1-\beta)\,(1 - G_1'(\rho))} \end{aligned}$$

$$= 1 + \frac{G_0'(\rho)\rho}{(1-\beta)\left(1-G_1'(\rho)\right)}$$
$$= 1 + \frac{\langle k\rangle \rho^2}{(1-\beta)\left(1-G_1'(\rho)\right)}. \tag{4.35}$$

Here we have used the fact that $H_1(1) = \rho$ and $H_0(1) = 1 - \beta$ (as in (4.33)) to move from the first line to the second, and both (4.31) and (4.33) to move from the second to third. In the final line we have made use of the fact that

$$G_0'(\rho) = \sum_{k=0}^{\infty} k p_k \rho^{k-1} = \langle k\rangle \sum_{k=0}^{\infty} \frac{k p_k}{\langle k\rangle} \rho^{k-1} = \langle k\rangle \rho,$$

with the final equality above coming directly from (4.32). Note that (4.35) agrees with our previously found (4.18) precisely when $S = 0$ and $\rho = 1$, that is, when there is no giant component and extinction is certain. We now present Theorem 3.11 as an example using this new framework.

Example 4.17. *Let $\mathcal{G}_{n,p}$ be a binomial random graph with $np \to \lambda$, where λ is a constant. Then we know from Example 3.6 that in the limit, the degree distribution has generating function given by $G_0(z) = e^{\lambda(z-1)}$. Moreover, we mentioned after Definition 4.11 that when $np \to \lambda$, the forward degree distribution and the degree distribution are equal for $\mathcal{G}_{n,p}$. Hence we find that*

$$\rho = G_1(\rho) = G_0(\rho) = e^{\lambda(\rho-1)}$$

is solved by (using (4.33))

$$\beta = 1 - G_0(\rho) = 1 - \rho$$

where β satisfies

$$\beta + e^{-\beta\lambda} = 1, \tag{4.36}$$

a familiar result indeed. Moreover, noting that

$$G_1'(\rho) = \lambda e^{\lambda(z-1)} = \lambda\rho, \tag{4.37}$$

we can apply (4.35) to find the average component size in $\mathcal{G}_{n,p}$. Since $np \to \lambda$, we have that $\langle k\rangle \to \lambda$, and hence as $n \to \infty$, the average component size in $\mathcal{G}_{n,p}$ is

given by

$$\begin{aligned}\langle s\rangle &= 1 + \frac{\lambda\rho^2}{(1-\beta)(1-\lambda\rho)} \\ &= 1 + \frac{\lambda\rho^2}{\rho(1-\lambda\rho)} \\ &= \frac{1}{1-\lambda(1-\beta)}. \end{aligned} \tag{4.38}$$

This is a result that we did not previously calculate in Chapter 3.

This is in fact a well known result [9], and similar results can be calculated for other common distributions. It is worth mentioning that various other quantities of interest can be calculated using this framework. In particular, we can find closed form expressions for the asymptotic number of t-th neighbours, the average path length, and even describe the behaviour of $H_0(z)$ close to the phase transition. Aside from its stark simplicity, a nice feature of the above framework is that it translates seamlessly into modelling diffusive processes. We discuss an application of our results from this section in Chapter 5.

4.3.7 Limitations and Extensions

We briefly mention here a few of the limitations of the analysis in this chapter and the extensions which have been made to the results of Molloy & Reed [42, 43]. Firstly, the canonical ensemble loses some of the generality of the microcanonical ensemble, though for practical purposes and modelling it is easier to work with. The original result of Molloy & Reed concerned sequences $\{\mathbf{d}_n\}_{n\in\mathbb{N}}$ of degree sequences $\mathbf{d}_n = (d_0(n), d_1(n), \dots)$ with $d_i(n) = 0$ for all $i \geq n$ and $\sum_{i\geq 0} d_i(n) = 2m$. They proved that when these degree sequences converge "nicely" to a distribution $\mathcal{D}$, then the threshold $\sum_k k(k-2)p_k = 0$ (as in (4.20)) for the emergence of the giant component held under two conditions. Firstly, they required that $d_i(n)$ had to be "sparse", meaning that

$$\sum_k k d_i(n) = nK(1 + o(1))$$

for some constant K. Secondly we must have $d_i(n) = 0$ whenever $i > n^{1/4-\epsilon}$. This final condition is really the most restrictive of all of the conditions imposed by Molloy & Reed in [42], as it allows a maximum degree of at most $n^{1/4-\epsilon}$. Molloy & Reed highlight in their results that the threshold for the emergence of the giant

component does not behave nicely for certain degree sequences with maximum degrees above $n^{1/4-\epsilon}$. For example, consider the degree sequence with $d_1(n) = n - \lceil n^{0.9} \rceil$, $d_i(n) = \lceil n^{0.9} \rceil$ if $i = \lceil \sqrt{n} \rceil$, and $d_i(n) = 0$ otherwise. Then it is easy to see (using notation from Lemma 4.9) that $n_1(\mathbf{d})/n \to 1$ and $n_i(\mathbf{d})/n \to 0$ for $i > 1$ as $n \to \infty$. Hence the threshold equation gives

$$\sum_k k(k-2)p_k = -1 < 0.$$

However, one can show that there are enough vertices of degree $\sqrt{n}$ to ensure that a giant component containing $n - o(n)$ vertices exists w.h.p. It is clear that at the time of [42, 43], the threshold had not been fully understood nor characterised; at least when $\mathbf{d}$ is not sufficiently sparse.

Bollobás & Riordan [10] (among others, for example [21, 31, 33]) impose significantly weaker restrictions on the degree sequence $\mathbf{d}_n$ for the emergence of the giant component. (4.23) implies that the maximum degree of a graph generated within the canonical ensemble is $o(n)$. In fact, Equations (4.22) and (4.23) are the only restrictions imposed by Bollobás & Riordan, though they are stated in terms of convergence of a given (fixed) degree sequence $\{\mathbf{d}_n\}_{n\in\mathbb{N}}$ of degree sequences $\mathbf{d}_n$, rather than for a random degree sequence $\mathbf{d}$. Indeed, in [10], the six or so restrictions of Molloy & Reed were replaced by these two simple convergence conditions, though these still do not fully characterise the emergence of the giant component. For example, the so called "power-law", and other networks with heavy tailed distributions, which we mentioned earlier in Chapter 4 are not covered by the results of either Molloy & Reed or Bollobás & Riordan.

Joos, Peranau, Rautenbach & Reed [33], settled the question of finding a threshold function for the emergence of the giant component for *arbitrary* degree sequences. Consider the following example from their paper [33].

Example 4.18 (Joos et al. [33])**.** *Let $n = k^2$, where k is a large odd integer. Take $d_1(n) = d_2(n) = \cdots = d_{n-1}(n) = 1$, and $d_n(n) = 2k$. Then*

$$\frac{\sum_{k=1}^n d_k(d_k-2)}{\sum_{k=1}^n d_k} = \frac{4k^2 - 4k - (n-1)}{2k+n-1} \approx 3,$$

and so the Bollobás & Riordan or Molloy & Reed approach would both suggest that with probability $1 - o(1)$ *there is a giant component. However, with probability* 1 *any graph* G *on the above degree distribution is the disjoint union of a star with* $2k$ *leaves and* $\frac{n-2k-1}{2}$ *components of order* 2 *and hence it has no giant component.*

The threshold condition imposed by [33] is based on looking at how long one expects to find "new" vertices in a BRP on a given graph. They explain,

> Intuitively, since the probability that we explore a vertex is essentially proportional to its degree, in lower bounding the length of the period during which the expected increase remains positive, we could assume that the exploration process picks at each step a highest degree vertex that has not been explored yet. Moreover, note that vertices of degree 2 have a neutral role in the exploration process as exposing such a vertex does not change the number of open edges, provided we assume that our component locally looks like a tree (which turns out to be a good approximation around the critical window).

The statement of the condition for the emergence of the giant component is rather technical and we do not provide it here. We suggest that an area for further research would be a more intuitive and equivalent criterion which still holds for an arbitrary degree distribution. It may be useful to have a criterion which also improves on the computational feasibility so that one may more easily check whether a graph on a given degree distribution yields a giant component w.h.p..

Recall that the reason we wanted to study the giant component for arbitrary degree distributions was because the degree distribution is an important characteristic of real world networks that is not captured by $\mathcal{G}_{n,p}$. We turn now to an application of the methods we have presented in this chapter to the field of Game Theory, an important area in the study of human behaviour. This field has been studied by economists, psychologists, computer scientists, and philosophers alike. We focus on an economic approach in what follows.

Chapter 5

Information Cascades and Diffusion Games

In this chapter, we analyse a model of diffusion on networks and discuss the economic implications of our results.

We discussed many of the properties exhibited by real world networks at the beginning of Chapter 4. Random graphs have become a central technique in modelling complex networks (see [28, Chapters 6 - 8] for a detailed treatment). In the case where one has data on a network, in particular, if one knows the degree $\boldsymbol{d}_i$ of each member in a network, the Configuration Model (Section 4.2) serves as a good benchmark to test whether or not the observed network exhibits any additional structure than a network chosen uniformly at random over the simple graphs with the degree sequence $(\mathbf{d}_1, \ldots, \mathbf{d}_n)$. In economics models, a variable is called *exogenous* if it imposed on the model, rather than being determined by the model. In what follows, we will take the network structure as exogenous, determined by a degree sequence from which the configuration model generates a graph, as in Section 4.2.

The process of the spread of information (gossip, disease) over a network is called *diffusion*. When looking at diffusion on random graphs, one is usually interested in when an *epidemic* or *contagion* can occur. An epidemic is defined to be a connected set of infected vertices which makes up constant fraction of all the vertices in the graph. In other words, an epidemic is a giant component of "infected" individuals. In the economics literature, an epidemic is often called a *large cascade*. We will use this language in Section 5.1.

The model of diffusion of information developed by Watts [55] has recently been generalised by Sadler [50]. In this chapter, we present Sadler's model for "single-

type diffusion games", as in [50]. An original finding of Sadler [50] is that when information has spread to a strategic player, that player learns something about their position in the network. This may not be immediately clear, but we will go into further detail on this in Section 5.2.2. Sadler calls this effect *viral inference*. In the conclusion to the thesis, we present some original remarks suggesting how one might generalise Sadler's results for the single-type diffusion game, to the case where individuals have some information about the time at which the they are "exposed". Sadler's paper is a marriage of the fields of probabilistic combinatorics and game theory, and as such it uses random graphs to provide economic insights.

5.1 Single-Type Diffusion Games

We follow the conventions and notation of Sadler [50] throughout this section. There are quite a few definitions required to set up Sadler's model, though some of them will be familiar from our work in Chapter 4.

A diffusion game is a sequence of n-player games $\{\Gamma^{(n)}\}_{n\in\mathbb{N}}$, where for each $n \in \mathbb{N}$, the game $\Gamma^{(n)} = (\mathbf{d}^{(n)}, S, V, u)$ is the four-tuple made up of the following components.

1. The vector $\mathbf{d}^{(n)} = (d_1^{(n)}, \ldots, d_n^{(n)})$ is the *degree sequence* for the game.

2. The set $S = \{0, 1\}$ describes the possible actions an player can take. If player i chooses the action $s_i \in S$ such that $s_i = 1$, we say that player i is an *adopter*.

3. Player's private "values" for adoption are drawn from the probability distribution V on $[0, 1]$. Note that in this chapter, V will denote a probability distribution, *not* the vertex set $[n]$.

4. The payoff to an individual from adoption is determined by the utility function $u(v, d, a)\colon [0, 1] \times \mathbb{N}^2 \to \mathbb{R}$. The utility function depends on the private value, the degree of the individual, and the number of neighbours who adopt the product.

The idea that an individual's payoff from adoption depends on how many of their friends have adopted is one of the main features of externalities in binary decision problems which Watts [56] sought to address. Note that the utility function is the same for all individuals, but the exact payoff will vary depending on v, d and a.

Following Sadler, we assume that u is strictly increasing in v, and weakly increasing in a.

We note here that we will be working in the *microcanonical ensemble* in this chapter, so each degree sequence $\mathbf{d}^{(n)}$ is fixed, rather than $d_1^{(n)}, \ldots, d_n^{(n)}$ being drawn independently from some distribution $\mathcal{D}_n$. As such, we will need to specify the limiting properties of $\mathbf{d}^{(n)}$ so that $\lim_{n\to\infty} \mathbf{d}^{(n)}$ is well-behaved. As in Lemma 4.9, let $n_k(\mathbf{d}^{(n)})$ be the number of vertices of degree k in $\mathbf{d}^{(n)}$, and $m(\mathbf{d}^{(n)})$ the number of edges in a graph with degree sequence $\mathbf{d}^{(n)}$. (Note that because we are in the microcanonical ensemble, n_k and m are *not* random variables, though they were in Lemma 4.9). Then in order for $\lim_{n\to\infty} \mathbf{d}^{(n)}$ to be well-behaved, we assume that there exists a distribution $\mathcal{D}$ with finite expectation and with p.m.f. $\{p_k\}_{k\in\mathbb{N}}$, such that for each $k \in \mathbb{N}$,

$$\lim_{n\to\infty} \frac{n_k(\mathbf{d}^{(n)})}{n} = p_k, \text{ and} \tag{5.1}$$

$$\lim_{n\to\infty} \frac{m(\mathbf{d}^{(n)})}{n} = \frac{\mathbb{E}(\mathcal{D})}{2}. \tag{5.2}$$

Note that Lemma 4.9 proves precisely that Equations (5.1) and (5.2) hold in the canonical ensemble. We now continue to describe the timing of the game.

The game $\Gamma^{(n)}$ takes place over $n+1$ time periods $t = 0, 1, \ldots, n$. Write $s_i(t)$ for the action of player i at time t. We now describe the timing of the game. At $t = 0$, every player has action $s_i(0) = 0$; that is, no player has adopted. We now introduce a player called "nature" who will act at $t = 1$. This is a common technique in the study of dynamic games of incomplete information (of which the single-type diffusion game is one). At $t = 1$, nature makes three moves.

1. First, nature draws a graph $G = \mathbb{G}_{\mathbf{d}} \in \mathscr{G}_{\mathbf{d}}$ uniformly at random from the set of all (simple) graphs with degree sequence $\mathbf{d}^{(n)}$. Each vertex of G represents a player in the game. We will use the term *network* and *graph* interchangeably when referring to G.

2. Second, nature draws a value $v_i \in [0, 1]$ independently from V for each player i.

3. Finally, nature chooses a "seed" (a first-adopter) uniformly at random. The seed switches to action 1 at time $t = 1$.

One can think of the "seed" as the inventor of a product who wants to encourage others to adopt their invention. Alternatively, one can think of the seed as an

individual who manufactures a piece of gossip which they then tell to all of their friends. Modelling infectious diseases does not really fit into this framework since agents are strategic and the adoption decision is endogenous (that is, decided by each player).

Let G_i denote the set of players who are neighbours of i in G. Player i acts at time $t_i := 1 + \min\{t\colon \text{there exists } j \in G_i \text{ with } s_j(t) = 1\}$. We say that player i is *exposed* at time t_i, since this is the first time at which a neighbour of i has adopted. At this point, player i makes a once-and-for-all decision as to whether or not they will adopt. Note that it is possible for player i to never be exposed, in which case they will have $s_i(t) = 0$ for all t. Define $s_i = s_i(n)$ to be the final action of player i, and define $a_i = \sum_{j \in G_i} s_j$, the number of i's neighbours who adopt. If $s_i = 1$, then player i earns the payoff $u(v_i, d_i, a_i)$. If $s = 0$, then player i earns zero (in economics, this is called *normalising the outside option*).

The information structure of the game is as follows. All players know the degree sequence $\mathbf{d}^{(n)}$, and the value distribution V. Up until the point at which a player is given the opportunity to act (if at all), they have only two pieces of private information: their degree d_i, and their value v_i. If player i gets exposed, then they can infer that at least one of their neighbours has adopted, but they do not know how many, nor which neighbour. Moreover, they do not know the time t_i at which they are exposed. This is an assumption which we discuss how one might relax in Section 5.2.3. After player i has been exposed, they also know whether or not they chose to adopt, that is, whether $s_i(t_i) = 1$. At time t_{i+1}, each player $j \in G_i$ will infer that at least one of their neighbours has adopted, so j will assign some positive probability to the event that i has adopted.

We remark here that the games outlined in this section are called "single-type" diffusion games, because Sadler [50] generalises these games to allow for multiple "types" of players. The analysis of multi-type diffusion games is based primarily on a generalisation of uniform branching processes which allow a finite set of possible offspring distributions at a given step in the branching process. These are called multi-type branching processes, and are a special case of the non-uniform branching process which we introduced in Remark 3.7 and discussed further in Section 3.1.4. A rigorous treatment of multi-type branching processes can be found in [3, Chapter 5].

To "solve" a diffusion game, we need to introduce an *equilibrium* concept. Equilibria are a fundamental idea in game theory. They describe a "steady state" of the game, in which all individuals are playing in such a way that any deviation from the equilibrium strategy could not result in an individual receiving a higher (expected) payoff. We first introduce the notion of a strategy profile.

Definition 5.1. A *strategy profile* is a function $\sigma(v, d)\colon [0,1]\times\mathbb{N} \to S$, which maps value-degree pairs to adoption decisions.

If $\sigma(v_i, d_i) = 1$, then player i switches to action 1 if given the opportunity. Sadler calls such players *potential adopters*. If player i is never exposed, or if $\sigma(v_i, d_i) = 0$, then player i never adopts. For a player with degree d, write A_d for the (random) number of neighbours who are potential adopters. The distribution of A_d will depend on the strategy profile σ, and will be the same for all players with $d_i = d$.

If we fix a strategy profile σ, then since player i knows both v_i and d_i, a *belief* about the distribution of A_{d_i} is a sufficient statistic for player i to be able to maximise their expected payoff. This is because player i's payoff depends only on v_i, d_i, and a_i, where the distribution of a_i given d_i is precisely A_{d_i}. Player i forms beliefs about the distribution A_d at time t_i, that is, the moment that they are exposed. Under Sadler's model, player i does not know t_i, and therefore cannot use t_i in the formation of their belief about the distribution of A_d. We now introduce the equilibrium concept which will be used throughout the chapter. Write $\mathbb{E}_\sigma^{(n)}$ for an expectation taken in $\Gamma^{(n)}$ assuming players follow the strategy profile σ.

Definition 5.2. Let $\sigma' = \sigma'(v, d)$ and $\sigma = \sigma(v, d)$ be strategy profiles. We say that σ' is a *limit best-reply* to σ, if

$$\lim_{n\to\infty} \mathbb{E}_\sigma^{(n)}\left(u(v, d, A_d)\right) \geq 0 \text{ whenever } \sigma'(v, d) = 1, \text{ and} \tag{5.3}$$

$$\lim_{n\to\infty} \mathbb{E}_\sigma^{(n)}\left(u(v, d, A_d)\right) \leq 0 \text{ whenever } \sigma'(v, d) = 0. \tag{5.4}$$

The strategy profile σ is a *limit equilibrium* if σ is a limit best-reply to itself.

If Equations (5.3) and (5.4) hold for some finite n, then σ is called a *Perfect Bayesian Equilibrium* (this is a standard concept in game theory, see [23] for more details). It is an immediate result of standard fixed point theorems that a Perfect Bayesian

Equilibrium exists in $\Gamma^{(n)}$ for all $n \in \mathbb{N}$, and also that a limit equilibrium exists. Proofs of these statements can be found in [50, Propositions 1 - 2]. Now that we have defined the single-type diffusion game, we describe what kind of properties of the game we would like to analyse.

5.2 Analysis of Single-Type Diffusion Games

We are interested in analysing how many players adopt, and how long the process takes. In the game $\Gamma^{(n)}$, write $X_n(t)$ for the number of adopting players at time t, that is,

$$X_n(t) = \sum_{i=1}^{n} s_i^{(n)}(t).$$

Write $X_n = X_n(n)$ for the final number of adopters. Then the long-run fraction of players who adopt is given by the random variable

$$\alpha_n \coloneqq \frac{X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} s_i^{(n)}, \tag{5.5}$$

and the time it takes for a fraction x of these players to adopt is given by the random variable

$$\tau_n(x) \coloneqq \min \left\{ x \colon \frac{X_n(t)}{X_n} \geq x \right\}. \tag{5.6}$$

In Section 5.2.1, we will analyse the limiting properties of α_n. Sadler [50] also analyses the limiting behaviour of τ_n, but since his results depend on average path lengths, a topic which we did not cover in Chapter 4, we will focus only on α_n. An important factor in our analysis will be whether or not G has a giant component. In Section 5.2.2, we discuss how these outcomes feed into players' equilibrium beliefs.

5.2.1 Mapping Strategies to Outcomes

Suppose we take σ to be a fixed strategy profile. Then once a graph G has been drawn by nature, all potential adopters are determined by the values V. In other words, the n independent draws from V establish the *potential adopter network*, that is, the subgraph $H \subseteq G$ of all potential adopters. The location of the seed then determines who actually adopts. The number of adopters can be at most the size of the largest component in the potential adopter network.

Recall that we assume $\{\mathbf{d}^{(n)}\}_{n\in\mathbb{N}}$ converges to a distribution $\mathcal{D}$ with finite mean (see Equations (5.1) and (5.2)). Remaining consistent with our notation in Section 4.3.1, let $\mathcal{D}'$ be the *forward distribution* corresponding to $\mathcal{D}$. Let

$$G_0(z) = \sum_{k=0}^{\infty} p_k z^k$$

be the generating function for $\mathcal{D}$. Then by (4.9), the generating function for $\mathcal{D}'$ is given by

$$G_1(z) = \frac{1}{\langle k \rangle} G_0'(z).$$

We will also need notation to describe the probability that different "types" of players are potential adopters. Taking σ as fixed, let a_σ denote the probability that a player drawn at random is a potential adopter. Let $a_{\sigma,d}$ be the probability that a random player of degree d is a potential adopter. Finally, let b_σ denote the probability that a randomly chosen neighbour of a randomly chosen player is a potential adopter.

Definition 5.3. Fix a strategy profile σ.

1. The *player adoption probability* is $a_\sigma = \mathbb{E}_V \mathbb{E}_{\mathcal{D}} (\sigma(V, \mathcal{D}))$.
2. The *degree-d adoption probability* is $a_{\sigma,d} = \mathbb{E}_V (\sigma(V, d))$.
3. The *forward adoption probability* is $b_\sigma = \mathbb{E}_V \mathbb{E}_{\mathcal{D}'} (\sigma(V, 1 + \mathcal{D}'))$.

Here, $\mathbb{E}_V$ indicates that the expectation is taken over the distribution of V, and similarly for $\mathbb{E}_{\mathcal{D}}$ and $\mathbb{E}_{\mathcal{D}'}$.

Since adoption can only spread among the potential adopter network, the existence of a giant component will be dependent on the distribution of degrees in the potential adopter network. It is therefore important to define the degree distribution among potential adopters.

Definition 5.4. *Fix a strategy profile σ. The* potential adopter degree distribution *$\mathcal{D}_\sigma$ is defined by*

$$\mathbb{P}(\mathcal{D}_\sigma = k) = \frac{a_{\sigma,k}\, p_k}{\sum_{j=0}^{\infty} a_{\sigma,j}\, p_j}, \tag{5.7}$$

where $p_k = \mathbb{P}(\mathcal{D} = k)$.

An *epidemic* or *large cascade* of information can only be possible when there is a giant component in the network of potential adopters. In order to analyse the existence of a giant component of potential adopters, we follow a similar idea to Section 4.3.2. Let $G_{0,\sigma}(z)$ be the generating function for $\mathcal{D}_\sigma$, and $G_{1,\sigma}$ be the generating function for the forward degree distribution corresponding to $\mathcal{D}_\sigma$. Then we have that

$$G_{1,\sigma}(z) = \frac{G'_{0,\sigma}(z)}{G'_{0,\sigma}(1)}. \tag{5.8}$$

Assume that the potential adopter network is "tree-like". Then we can approximate this network by a branching process. Some of the following details were absent from Sadler [50], but I provide them here using ideas from Watts [55].

What is the offspring distribution of this branching process? To answer this, consider choosing at random a player i from the network. The probability that i is an adopter of degree k is given by $\mathbb{P}(\mathcal{D}_\sigma = k)$. Now consider randomly follow one of i's neighbours to another player. The probability that this neighbour is a potential adopter is b_σ. That is, with probability $1 - b_\sigma$, a randomly chosen neighbour of i is not a potential adopter. Consider the probability that a randomly chosen player is a potential adopter of degree k, and a randomly chosen neighbour of that player is a potential adopter. This defines the p.m.f. for the *forward "adoption" degree distribution*. By applying Theorem 2.22 to (5.8), the generating function for the forward adoption degree distribution is given by

$$G_{2,\sigma}(z) \coloneqq G_{1,\sigma}(1 - b_\sigma + b_\sigma z) = \frac{G'_{0,\sigma}(1 - b_\sigma + b_\sigma z)}{G'_{0,\sigma}(1)}. \tag{5.9}$$

It follows that the *forward extinction probability* ρ_σ for the generating function (5.9) is the smallest solution in $[0, 1]$ to the equation

$$G'_{0,\sigma}(1)\rho_\sigma = G'_{0,\sigma}(1 - b_\sigma + b_\sigma \rho_\sigma). \tag{5.10}$$

This is equivalent to [50, Proposition 3].

The argument which we used to derive the threshold for the emergence of the giant component in Section 4.3.3 is still valid here. It follows from (4.16) that a giant component of potential adopters exists if and only if

$$G'_{2,\sigma}(1) > 1. \tag{5.11}$$

By the chain rule, we have that

$$\left[\frac{d}{dz}G_{2,\sigma}(z)\right]_{z=1} = b_\sigma \frac{G''_{0,\sigma}(1)}{G'_{0,\sigma}(1)}. \tag{5.12}$$

Therefore, a giant component of potential adopters exists if and only if

$$b_\sigma \sum_{k=0}^{\infty} k(k-1)\,\mathbb{P}(\mathcal{D}_\sigma = k) > \sum_{k=0}^{\infty} k\,\mathbb{P}(\mathcal{D}_\sigma = k), \tag{5.13}$$

which, by substituting in (5.7), is equivalent to

$$\sum_{k=0}^{\infty} a_{\sigma,k} k(b_\sigma(k-1)-1)p_k > 0. \tag{5.14}$$

It is remarkable that we find an expression so similar to (4.20) given the added complexity of the model here. One can see that the left hand side of (5.13) has been "reweighted" by the probability that a randomly chosen neighbour is a potential adopter. This is because we are concerned with branching processes beginning at a randomly chosen vertex which spread only through potential adopters.

Now that we know when giant components of potential adopters will exist, we turn to the question of when large cascades will occur. Let $\mathcal{C}$ denote the largest connected component of potential adopters. Using the forward extinction probability ρ_σ, we can calculate the fraction of players with a connection to $\mathcal{C}$, and the fraction of players contained in $\mathcal{C}$. Following Sadler's terminology, we denote these by ζ_σ and ϕ_σ respectively.

Definition 5.5.

1. The *player diffusivity* is $\zeta_\sigma = 1 - G_0(1 - b_\sigma + b_\sigma\rho_\sigma)$.

2. The *potential adopter diffusivity* is $\phi_\sigma = p_\sigma\,(1 - G_{1,\sigma}(1 - b_\sigma + b_\sigma\rho_\sigma))$.

The player diffusivity ζ_σ is to the probability of triggering a large cascade, that is, the probability that the randomly chosen seed is connected to the giant component of potential adopters. The potential adopter diffusivity ϕ_σ is the fraction of vertices in the cascade.

Note that by definition of the extinction probability ρ_σ from (5.10), we have that

$$G_{1,\sigma}(1 - b_\sigma + b_\sigma \rho_\sigma) = \rho_\sigma$$

and hence $\phi_\sigma = 1 - \rho_\sigma$. Define α to be the random variable on outcomes $\{0, \phi_\sigma\}$ with probabilities

$$\mathbb{P}(\alpha = 0) = 1 - \zeta_\sigma, \quad \mathbb{P}(\alpha = \phi_\sigma) = \zeta_\sigma. \tag{5.15}$$

Then α describes the extent of diffusion in large networks. Recall α_n from (5.5), the long-run fraction of players who adopt. We now present a theorem which relates α_n to α.

Theorem 5.6 ([50, Theorem 1])**.** *Fix a profile σ. The fraction of players α_n who adopt converges in distribution to α.*

Proof. By Equations (5.1) and (5.2), the degree of a random potential adopter converges in distribution to $\mathcal{D}_\sigma$, and the probability that a random neighbour is a potential adopter converges to b_σ. We know from our analysis in Section 4.3.6 that the size of the giant component is given by the extinction probability of a branching process on $X_1 \sim \mathcal{D}_\sigma$ and $X_i \sim \mathcal{D}'_\sigma$ for all $i \geq 2$. To make this precise, one can use a concentration result analogous to Corollary 4.15 but for the microcanonical ensemble (see [10, Theorem 2] for a sufficient result).

Since the extinction probability of the forward adoption distribution $\mathcal{D}'_\sigma$ is given by ρ_σ, it follows that the extinction probability of a branching process on $X_1, X_2, \ldots,$ is

$$1 - p_\sigma + p_\sigma \rho_\sigma = 1 - \phi_\sigma, \tag{5.16}$$

as required. Here, $1 - p_\sigma$ represents the probability that the first vertex in the branching process is not a potential adopter. If they *are* a potential adopter, then the branching process goes extinct with probability ρ_σ, since each subsequent step in the exploration follows the forward adoption distribution $\mathcal{D}'_\sigma$.

Moreover, the probability that a randomly chosen seed *fails* to connect to the giant component is simply $1-\zeta_\sigma$. In this case, no large cascade is triggered and so $\alpha_n = 0$. □

Intuitively, Theorem 5.6 says that if the initial seed touches the giant component of potential adopters, then everyone in that component adopts. Otherwise, we get almost no adoption. Even if the initial seed connects to some component of potential adopters other than the giant component, this component is negligible in size as $n \to \infty$. The average number of new potential adopters at each step is given by $\nu_\sigma := b_\sigma \, \mathbb{E}(\mathcal{D}'_\sigma)$. The parameter ν_σ describes the rate at which adoption spreads. Sadler [50] calls ν_σ the *virality*. We now use the idea of virality to discuss limiting beliefs under different strategy profiles σ.

5.2.2 Limit Beliefs and Viral Inference

We argued in Section 5.1 that A_d, the random number of neighbours who are potential adopters for a player with degree d, is a sufficient statistic for players to maximise their expected utility. Here, we look at the limiting distribution of A_d. Following Sadler, we say that a strategy profile σ is *viral* if the potential adopter network contains a giant component, otherwise we say that σ is *non-viral*.

Before presenting theorems on the limiting distribution of A_d, we provide some intuition as to why the distinction between viral and non-viral equilibria is important in constructing beliefs. First, in economics we use the Latin term ex-ante, meaning "before the event", to refer to probabilities calculated by any individual before they receive any private information. In a non-viral equilibria, all connected components of potential adopters are finite and of similar size. Conditional on exposure, an individual knows that at least one of their neighbours has adopted. If σ is viral however, the ex-ante probability of exposure is bounded away from zero, and conditional on a player's exposure, w.h.p. that player is connected to the giant component. The following theorem describes the limiting distribution of A_d for a non-viral strategy profile σ.

Theorem 5.7 ([50, Theorem 3])**.** *Let σ be a non-viral strategy profile. As $n \to \infty$, the number of adopting neighbours A_d is distributed as $1 + \mathrm{Bin}(d-1, b_\sigma)$.*

Proof. If σ is non-viral, then by Corollary 4.13, any finite neighbourhood of a vertex is a tree. Hence in the diffusion process, a player can be exposed by at most 1 other player. This implies that exposure resolves uncertainty for exactly one neighbour. Moreover, any one player's adoption is independent of of whether or not their neighbours are also potential adopters. Therefore, if a player is exposed,

the $d-1$ neighbours who did not expose them are potential adopters with probability b_σ. It follows that $A_d \sim 1 + \text{Bin}(d-1, b_\sigma)$. □

The more interesting case is when σ is viral. In this case, exposure provides information about a player's position in the network. As $n \to \infty$, the probability that a random seed is in any non-giant component $o(1)$. Therefore, conditional on exposure in a viral equilibrium, a player is connected to the giant component with probability $1 - o(1)$.

Theorem 5.8 ([50, Theorem 3]). *Let σ be a viral strategy profile. As $n \to \infty$, the number of adopting neighbours A_d satisfies*

$$\mathbb{P}(A_d = k) = \frac{1 - \rho_\sigma^k}{1 - (1 - b_\sigma + b_\sigma \rho_\sigma)^d} \mathbb{P}(\text{Bin}(d, b_\sigma) = k). \tag{5.17}$$

Proof. Consider a BRP beginning at a potential adopter. This process goes extinct with probability ρ_σ. Since σ is viral, we have that $\rho_\sigma < 1$. Therefore, the probability that a randomly chosen potential adopter connects to the giant component of potential adopters is $1 - \rho_\sigma$. For a randomly chosen player i, let $\mathcal{S}_i$ denote the event that at least one of i's neighbours is a potential adopter who is connected to the giant component. The ex-ante probability that a neighbour who is a potential adopter fails to connect to the giant component is simply ρ_σ, independently for each neighbour. Therefore, since the probability that any neighbour of i is connected to the giant component is independent of any other neighbours, the probability that $\mathcal{S}_i$ does not happen is $(1 - b_\sigma + b_\sigma \rho_\sigma)^d$. This is similar to the argument we made in deriving (5.16). Hence conditional on exposure, we have

$$\begin{aligned} \mathbb{P}(A_d = k \mid \mathcal{S}) &= \frac{\mathbb{P}(\mathcal{S} \mid A_d = k) \cdot \mathbb{P}(A_d = k)}{\mathbb{P}(\mathcal{S})} \\ &= \frac{(1 - \rho_\sigma^k)\mathbb{P}(A_d = k)}{1 - (1 - b_\sigma + b_\sigma \rho_\sigma)^d}. \end{aligned}$$

Since $A_d \sim \text{Bin}(d, b_\sigma)$ ex-ante, this completes the proof. □

Interestingly, in a viral equilibrium, players can infer less from exposure. This is because there was already some positive probability that they were going to become exposed. As we mentioned in the introduction to this chapter, Sadler calls this effect *viral inference*.

5.2.3 Viral inference with knowledge of time

We conclude this chapter with some original suggestions for how one might generalise Sadler's results about viral inference, to the case where individuals have information about the time t_i at which they are exposed. This is a topic on which Sadler makes some remarks in [50, Section 6], though no suggestions are provided as to how one might proceed. Recall that we assumed in our description of the single-type diffusion game that individuals only have two pieces of private information upon exposure. Namely, an individual i knows d_i and v_i. In a non-viral equilibrium, knowledge of t_i upon exposure will have no effect on an individual's belief about A_d. In a viral equilibrium, a player who is exposed early on knows that the information has not had much time to spread, whereas a player exposed much later is almost certain that they are connected to the giant component of potential adopters.

Now suppose that every individual i knows the time t_i at which they are exposed. There is a very nice result by Newman [46] which derives the derives an explicit expression for the distribution of component sizes in the configuration model. The paper [46] uses similar methods to [47], though there are more subtleties associated with the derivation. Consider a branching process with offspring distribution $\mathcal{D}'_\sigma$. The expected offspring after n generations (or time periods) is

$$f_n \coloneqq 1 + \mathbb{E}(\mathcal{D}'_\sigma) + \mathbb{E}(\mathcal{D}'_\sigma)^2 + \cdots + \mathbb{E}(\mathcal{D}'_\sigma)^{n-1} = \frac{\mathbb{E}(\mathcal{D}'_\sigma)^n - 1}{\mathbb{E}(\mathcal{D}'_\sigma) - 1}, \tag{5.18}$$

assuming that $\mathbb{E}(\mathcal{D}'_\sigma) \neq 1$. Then an individual who finds out that they were exposed at time t_i infers that (roughly) f_{t_i} offspring have been born so far, and therefore that they are in a component of at least this size. Hence by using Newman's result [46], an exposed player can condition their belief about the distribution of A_d on the probability that they are in a component of size at least f_{t_i}. For certain models of random graphs (for example, $\mathcal{G}_{n,p}$) one could certainly get a closed form expression for this. Whether or not a "clean" explicit expression is possible for random graphs with an arbitrary degree distribution is still an open question.

I would conjecture that an explicit approximate expression which is correct in the limit is attainable. An approximate solution to this problem could be attained by allowing each player i to condition their belief about the distribution of A_d on the probability that a BRP beginning at a randomly chosen vertex goes extinct given that it has already produced t_i generations. This extinction probability will be

lower than the ex-ante extinction probability, since there may be more than one individual producing offspring in generation t_i.

If obtaining a closed form solution proves to be too difficult an exercise, we suggest giving players only partial information about the time at which they are exposed. One way to do this would be to introduce a *time threshold* $t = t(n) > 1$, such that upon player i's exposure, i knows whether $t_i \geq t(n)$, or $t_i \leq t(n)$. Practically, one could think of this as each player passing on an additional piece of information in the diffusion process: whether the object begin diffused is "new" or "old". If for player i, both $t_i \geq t(n)$, and $t_i \leq t(n)$ then we find ourselves back at the original problem. To avoid this, we could insist that $t(n)$ is never an integer. Even if we allow $t(n)$ to be an integer, we suspect that the probability that $t_i = t(n)$ will be $o(1)$. Introducing a time threshold would reduce the difficulty of the problem since only two conditional probabilities would need to be calculated. Generalising Sadler's model [50] is certainly a useful area for future research, and we hope that our work here points researchers towards solving one important aspect of this problem.

Chapter 6

Concluding Remarks

In this thesis, we have explored the threshold for the emergence of the giant component in random graphs.

In Chapter 3, we followed the pioneering work of Erdős and Rényi [18, 19], proving that $p = \frac{1}{n}$ is a sharp threshold for the emergence of the giant component in the binomial random graph. In Chapter 4 we extended this result to multigraphs with an arbitrary degree sequence, finding an explicit threshold expression for emergence of the giant component which was first derived by Molloy & Reed [42, 43]. We provided a heuristic argument from [47], and made the argument rigorous using techniques from [10]. Moreover, we obtained a concentration result for the canonical ensemble. This concentration result implied that our results held also for simple graphs with an arbitrary degree sequence, such as the binomial random graph. Although the question of existence of the threshold for the emergence of the giant component has been settled for random graphs with an arbitrary degree sequence, the structure of these graphs in the critical window is far from a closed question, and would be an interesting area for future study.

Finally, in Chapter 5, we discussed an application at the frontier of research in models of information cascades: an economic phenomenon observed in the adoption of new behaviours, habits, and technologies. Following a new model developed by Sadler [50], we demonstrated how the methods introduced in Chapter 4 give insight into when information cascades are possible. We also discussed the role of viral inference, and highlighted some of the limitations of the model developed by Sadler in [50].

At the end of Chapter 5 we made some suggestions as to how one might go about generalising Sadler's model. This is an important area for future research, since individual's often have some idea of when a piece of information was first made available. Another important extension to the model would be *endogenising* the network structure. This simply means allowing players to have actions which change the network structure, for example, allowing players to "make new friends". Network formation models are something we briefly mentioned in the introduction, but were beyond the scope of this thesis. Sadler's model provides a rich repository of new research possibilities with important real world ramifications.

References

[1] W. Aiello, F. Chung, and L. Lu. A random graph model for power law graphs. *Experimental Mathematics*, 10(1):53–66, 2001.

[2] M. Akbarpour, S. Malladi, and A. Saberi. Diffusion, seeding, and the value of network information. *SSRN Electronic Journal*, 2017.

[3] K. Athreya and P. Ney. *Branching Processes*. Springer, Heidelberg, 1972.

[4] N. Bailey. *The mathematical theory of infectious diseases and its applications*. New York: Hafner, 1975.

[5] A. Banerjee, A. Chandrasekhar, E. Duflo, and M. Jackson. The diffusion of microfinance. *Science*, 341(6144):1236498–1236498, 2013.

[6] A. Banerjee, A. Chandrasekhar, E. Duflo, and M. Jackson. Using gossips to spread information: Theory and evidence from two randomized controlled trials. *The Review of Economic Studies*, 86(6):2453–2490, 2019.

[7] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[8] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1(4):311–316, 1980.

[9] B. Bollobás. *Random Graphs*. Cambridge University Press, 2001.

[10] B. Bollobás and O. Riordan. An old approach to the giant component problem. *Journal of Combinatorial Theory, Series B*, 113:236–260, 2015.

[11] B. Bollobás and A. Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.

[12] A. D. Broido and A. Clauset. Scale-free networks are rare. *Nature Communications*, 10(1), 2019.

[13] A. Chandrasekhar and M. Jackson. A network formation model based on subgraphs. 2016.

[14] A. Clauset, C. R. Shalizi, and M. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[15] R. Diestel. *Graph Theory*. Electronic library of mathematics. Springer, Berlin, 2006.

[16] K. Dietz and J. Heesterbeek. Daniel bernoulli's epidemiological model revisited. *Mathematical Biosciences*, 180(1-2):1–21, 2002.

[17] P. Erdős. Some remarks on the theory of graphs. *Bulletin of the American Mathematical Society*, 53(4):292–295, 1947.

[18] P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959.

[19] P. Erdős and A. Rényi. On the evolution of random graphs. In *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, volume 5, pages 17–61, 1960.

[20] I. Foppa. *A Historical Introduction to Mathematical Modeling of Infectious Diseases*. Elsevier Science, San Diego, 2016.

[21] N. Fountoulakis and B. Reed. A general critical condition for the emergence of a giant component in random graphs with given degrees. *Electronic Notes in Discrete Mathematics*, 34:639–645, 2009.

[22] A. Frieze and M. Karoński. *Introduction to Random Graphs*. Cambridge University Press, 2015.

[23] D. Fudenberg and J. Tirole. Perfect bayesian equilibrium and sequential equilibrium. *Journal of Economic Theory*, 53(2):236–260, 1991.

[24] E. Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

[25] B. Golub and M. Jackson. How Homophily Affects Diffusion and Learning in Networks. *arXiv e-prints*, page arXiv:0811.4013, Nov 2008.

[26] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, Oxford, 2001.

[27] Josef Hadar and William R. Russell. Rules for ordering uncertain prospects. *The American Economic Review*, 59(1):25–34, 1969.

[28] R. Hofstad. *Random Graphs and Complex Networks*, volume 1. Cambridge University Press, 2016.

[29] F. Hollander. Probability theory: The coupling method. Mathematical Institute, Leiden University, 2012.

[30] M. Jackson. A survey of models of network formation: Stability and efficiency. Technical report, University Library of Munich, Germany, 2003.

[31] S. Janson and M. Łuczak. A new approach to the giant component problem. *Random Structures and Algorithms*, 34(2):197–216, 2009.

[32] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. John Wiley & Sons, Hoboken, 2000.

[33] F. Joos, G. Perarnau, D. Rautenbach, and B. Reed. How to determine if a random graph with a fixed degree sequence has a giant component. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2016.

[34] M. Krivelevich and B. Sudakov. The phase transition in random graphs: A simple proof. *Random Structures & Algorithms*, 43(2):131–138, 2012.

[35] P. Lazarsfeld and R. Merton. Friendship as a social process: a substantive and methodological analysis. *Freedom and Control in Modern Society*, pages 18–66, 1954.

[36] J. Leskovec, L. Adamic, and B. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web*, 1(1):5–es, may 2007.

[37] I. Lobel, E. Sadler, and L. Varshney. Customer referral incentives and social media. *Management Science*, 63(10):3514–3529, 2017.

[38] T. Łuczak. Component behavior near the critical point of the random graph process. *Random Structures & Algorithms*, 1(3):287–310, 1990.

[39] B. McKay and N. Wormald. Uniform generation of random regular graphs of moderate degree. *Journal of Algorithms*, 11(1):52–67, 1990.

[40] M. McPherson, L. Smith-Lovin, and J. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

[41] A. Mehrabian. The giant component, a report for advanced random graph theory reading course. University of Waterloo, 2010, https://www.cs.mcgill.ca/ amehra13/Articles/giant_component.pdf.

[42] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–180, 1995.

[43] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combinatorics, Probability and Computing*, 7(3):295–305, 1998.

[44] J. Montgomery. Social networks and labor-market outcomes: Toward an economic analysis. *The American Economic Review*, 81(5):1408–1418, 1991.

[45] H. Moore. Cours d'économie politique. by vilfredo pareto, professeur à l'université de lausanne. vol. i. pp. 430. i896. vol. II. pp. 426. i897. lausanne: F. rouge. *The ANNALS of the American Academy of Political and Social Science*, 9(3):128–131, 1897.

[46] M. Newman. Component sizes in networks with arbitrary degree distributions.

[47] M. Newman, S. Strogatz, and D. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2), 2001.

[48] J.-P. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, and A.-L. Barabasi. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, 2007.

[49] P. Pin S. Currarini, M. Jackson. An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045, 2009.

[50] E. Sadler. Diffusion games. *Forthcoming, American Economic Review*, 2019.

[51] T. Schelling. Hockey helmets, concealed weapons, and daylight saving: A study of binary choices with externalities. *The Journal of Conflict Resolution*, 17(3):381–428, 1973.

[52] Q. Telesford, K. Joyce, S. Hayasaka, J. Burdette, and P. Laurienti. The ubiquity of small-world networks. *Brain Connectivity*, 1(5):367–375, 2011.

[53] F. Walsh. Superbugs to kill 'more than cancer' by 2050. *BBC News*, 2014, https://www.bbc.com/news/health-30416844.

[54] H. Watson and F. Galton. On the probability of the extinction of families. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 4:138–144, 1875.

[55] D. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002.

[56] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[57] N. Wormald. *Some problems in the enumeration of labelled graphs.* Ph.d. thesis, Newcastle University, 1978.

[58] Y. Zhang, E. Kolaczyk, and B. Spencer. Estimating network degree distributions under sampling: An inverse problem, with applications to monitoring social media networks. *The Annals of Applied Statistics*, 9(1):166–199, 2015.