University of Zurich
Institute of Mathematics

Master Thesis

Phase transition of Erdös-Rényi Graphs via Binomial and Poisson Branching Processes

Author:
Christian Tancredi

Supervisor:
Prof. Dr. Jean Bertoin

Co-Supervisor:
Daniela Portillo Del Valle

July 2026

Acknowledgements

At the end of this work, I would like to express my sincere gratitude to all those who, in different ways, have made this thesis possible.

I would like to thank my supervisor, Prof. Dr. Jean Bertoin, for accepting to supervise this thesis and for giving me the opportunity to work on such a fascinating topic at the intersection of random graphs and branching processes.

My deepest gratitude goes to my co-supervisor, Daniela Portillo Del Valle, who has truly been the guiding presence behind this entire work. She proposed the topic, suggested the references, and patiently followed every single step of the project, from the first hesitant attempts to the final revisions. Her detailed feedback, her sharp comments and the countless discussions we have had together have shaped not only this thesis but also, more broadly, my way of doing mathematics. For her patience, her availability and her constant care, I am sincerely and deeply grateful.

Muchísimas gracias por todo.

On a more personal note, I want to thank my friends, who have shared with me the joys and the struggles of these years and who have always been there when it mattered most. In particular, I wish to thank Matteo, Timo, Diana, Naty, Fabian, Fra, Pietro, Sabrina and Jona: the time spent together, both inside and outside the university, has made this whole journey infinitely lighter and more meaningful.

My deepest gratitude goes to my family. To my parents, Jeannette and Francesco, for their unconditional love, their unwavering support and the countless sacrifices that have allowed me to pursue my studies up to this point: nothing of what I have achieved would have been possible without you. To my brother, Alessandro, for being a constant presence in my life and a point of reference I have always been able to count on. Thank you for the laughs, the silences, the disagreements and the complicity that only a brother can give: knowing that you are there, no matter what, has made every difficult moment easier to face.

Finally, I wish to extend a heartfelt thank you to all those whose names do not appear explicitly in these pages but who have nonetheless been part of this journey: the friends, the colleagues and the people I have had the chance to meet along the way. Each of you has contributed, even in ways you may not realise, to the person I am today and to the work that follows. If your name is not written here, please know that it is only because the page is too short, and not because the gratitude is.

Grazie a tutti, di cuore.

Christian

Abstract

The Erdős–Rényi random graph $G(n,p)$ is the random graph on n vertices in which every pair of distinct vertices is connected, independently of all the others, by an edge with probability $p \in [0,1]$. Despite its simple definition, the model exhibits a phase transition in the structure of its connected component. A fundamental discovery, going back to the seminal work of Erdős and Rényi, is that, in the sparse regime $p = \lambda/n$, the size of the largest connected component of $G(n,p)$ undergoes a sharp phase transition at $\lambda = 1$: below this critical value all components are of logarithmic order, while above it a unique giant component containing a positive fraction of the vertices emerges.

In this thesis we study the phase transition of the Erdős–Rényi graph $G(n,p)$ from the perspective of Galton–Watson branching processes. By analysing the breadth–first exploration of connected components, we establish a precise comparison between component sizes in $G(n,p)$ and the total progeny of suitable branching processes. We use this comparison to study the subcritical, critical, and supercritical regimes. In particular, we show how extinction and survival of branching processes correspond to the absence or emergence of a giant component, and we discuss how branching process techniques can be used to derive quantitative results on component sizes. The exposition is based primarily on the framework developed by van der Hofstad, with additional context from the classical literature on random graphs.

Contents

1 Introduction **3**

2 Preliminaries **7**
2.1 Graph Theory Basics 7
2.2 Erdős–Rényi Random Graphs 11
2.2.1 A large-deviation constant for Poisson variables 12

3 Galton–Watson Branching Processes **15**
3.1 Definitions and first properties 16
3.2 Total progeny 21
3.2.1 Expected generation size 22
3.2.2 The critical case 24
3.2.3 Tail of the total progeny in the subcritical case 25
3.3 Poisson and binomial branching processes 26
3.3.1 Binomial branching processes 26
3.3.2 Poisson branching processes 27
3.3.3 From binomial to Poisson: the sparse limit 28
3.3.4 The conjugate parameter 29

4 Phase Transition in Erdős–Rényi Graphs **31**
4.1 Comparison with Branching Processes 33
4.2 Stochastic Domination of Cluster Sizes 35
4.2.1 Exploration process 35
4.2.2 Upper bound: domination by a branching process 38
4.2.3 Lower bound: domination from below 40
4.3 Subcritical Regime ($\lambda < 1$) 43
4.3.1 Expected component size 43
4.3.2 Size of the largest component 44
4.3.3 Interpretation 47
4.4 Supercritical Regime ($\lambda > 1$) 47
4.4.1 Strategy of the proof 48
4.4.2 Step 1: the cluster tail and the expected mass in large clusters 49
4.4.3 Step 2: concentration of the mass in large clusters 50
4.4.4 Step 3: no middle ground 51

4.4.5 Step 4: completion of the proof 53
4.4.6 Uniqueness of the giant component 54
4.4.7 The duality principle 54
4.4.8 Fluctuations of the giant component 56
4.5 Critical Regime and the Critical Window 56
4.5.1 Breakdown of the classical branching process approximation . 56
4.5.2 Size of the largest components 57
4.5.3 Scaling limits and exploration processes 58

5 Conclusions and Further Directions **61**

Appendix Overview **63**

A Probabilistic Background **65**

B Mathematical Tools **67**

List of symbols **69**

Bibliography **71**

Chapter 1

Introduction

This thesis studies the phase transition in the Erdős–Rényi random graph $G(n, \lambda/n)$ using Galton–Watson branching processes. Our main object of interest is the size of the largest connected component as $n \to \infty$.

A *graph* is, in its simplest form, a finite collection of points, called *vertices*, together with a list of pairs of points, called *edges*, which we think of as direct connections between the corresponding vertices. Two vertices are said to belong to the same *connected component* when one can be reached from the other by following a chain of edges, and the size of the largest such component, which we shall call the *largest connected component* and denote by $|\mathcal{C}_{\max}|$, is the central quantity of interest in this thesis. Such structures appear in a vast variety of contexts; social interactions, the World Wide Web, neural networks, transportation systems, biological pathways and a recurring feature of these real-world examples is that the precise list of edges is either too intricate to be specified deterministically, or genuinely random in nature. It is therefore natural to model the network itself as a random object and to ask which of its properties hold typically; this is the perspective that has driven the field of random graphs since its inception.

The simplest such probabilistic model was introduced independently by Erdős and Rényi [16, 17] and by Gilbert [20] at the end of the 1950s, and provides the canonical framework in which fundamental phenomena such as the appearance of thresholds, phase transitions, and the emergence of macroscopic clusters can be studied in a rigorous way. Their work showed that several graph properties undergo sharp transitions as the edge density varies. In particular, above a critical value, a component containing a positive fraction of the vertices emerges; the so-called *giant component*. The model itself is defined as follows: fix an integer $n \geq 1$ and a parameter $p \in [0, 1]$, called the *edge probability*; the *Erdős–Rényi random graph* $G(n, p)$ is the random graph on the vertex set $\{1, \dots, n\}$ obtained by including, independently for every pair of distinct vertices $i \neq j$, an edge between them with probability p and leaving them disconnected with probability $1 - p$. Equivalently, every one of the $\binom{n}{2}$ possible edges is present independently with probability p. The parameter p thus tunes the density of the network: small values of p produce a graph with few edges and many isolated parts, while values of p close to 1 produce

an almost complete graph. Despite its striking simplicity, $G(n,p)$ already exhibits a remarkably rich structure and has been the central object of a vast literature [9, 15, 19, 24]; throughout this thesis we shall mainly follow the modern, probabilistic exposition of van der Hofstad [41], which develops the theory using tools closely related to branching processes.

The basic question that we address is the following: when n is very large, how big is $|\mathcal{C}_{\max}|$ in $G(n,p)$? Already in their pioneering paper, Erdős and Rényi discovered that the answer depends crucially on the value of p, and that this dependence is far from continuous. The relevant scale on which interesting behaviour occurs is the so-called *sparse regime*, in which the average number of neighbours of any given vertex stays bounded as n grows. Since a fixed vertex of $G(n,p)$ is connected to each of the other $n-1$ vertices independently with probability p, its expected number of neighbours, which we call its *average degree*, equals $(n-1)p$; the natural choice is therefore

$$p \;=\; \frac{\lambda}{n}, \qquad \lambda > 0 \text{ fixed},$$

so that the average degree converges to λ and each vertex has, on average, roughly λ neighbours in the limit. The fundamental finding of Erdős and Rényi [17] is that, in this sparse regime, the typical size of $|\mathcal{C}_{\max}|$ in $G(n,\lambda/n)$ undergoes a sudden change at the value $\lambda = 1$. Below the threshold ($\lambda < 1$), all components are tiny, of size at most logarithmic in n, and the graph is fragmented in many small pieces. Above the threshold ($\lambda > 1$), one single component already contains a positive fraction of all vertices, while all other components are still of logarithmic order; this macroscopic component is precisely the giant component mentioned above. Exactly at the threshold ($\lambda = 1$), the largest components are of intermediate size, of order $n^{2/3}$, and exhibit heavy-tailed fluctuations [8, 39].

There are two complementary reasons for studying this phenomenon, one applied and one structural. From the applied side, the emergence of the giant component is the simplest mathematical incarnation of a property that one observes in many real-world networks: a slight increase in the density of connections may switch the network from being globally fragmented to being globally connected, and the Erdős–Rényi model has therefore become the canonical reference against which more realistic models of random networks are compared [15, 41]. From the structural, probabilistic side, the change of behaviour at $\lambda = 1$ is one of the cleanest and most studied examples of a *phase transition* in discrete probability, and it admits a transparent explanation in terms of a much older probabilistic model: the *Galton–Watson branching process*. The latter was introduced by Watson and Galton in 1875 [42] to study the extinction of family surnames over generations. In its simplest version, the process starts from a single ancestor; each individual of any given generation then produces, independently of all the others, a random number of children in the next generation, all distributed according to a common offspring law. The basic question is whether the population eventually dies out or survives forever. The celebrated answer is that this depends only on the mean number of children per individual: if this mean is at most 1 (and the offspring distribution is non-degenerate) extinction is certain, while if the mean exceeds 1 the population

survives with positive probability. Galton–Watson processes have since become a fundamental tool in probability and have found applications well beyond demography, in genetics, epidemiology and physics, among others [5, 22, 23, 27].

Breadth-first exploration gives the connection between Erdős–Rényi components and Galton–Watson processes. Starting from a vertex v, we successively reveal its unexplored neighbours. During the early stages of the exploration, the number of new neighbours is approximately distributed as Poisson(λ), so the exploration is well approximated by a Galton–Watson process with Poisson offspring distribution. Extinction corresponds to a small component, while positive survival probability is related to the emergence of a giant component.

This branching-process viewpoint goes back to the classical random-graph literature [8, 25]. In Chapter 4, following the approach of van der Hofstad [41], the comparison is made rigorous through stochastic domination by suitable Galton–Watson processes. The aim of this thesis is to develop this branching-process viewpoint in a systematic way and to use it to analyse the phase transition in the Erdős–Rényi random graph. In particular, we focus on the comparison between breadth-first exploration processes and Galton–Watson branching processes, which leads to precise stochastic domination results and sharp asymptotic descriptions of component sizes in the subcritical, critical, and supercritical regimes. The exposition follows mainly Chapters 3–5 of van der Hofstad [41], with additional input from the classical references [9, 15, 19, 24].

Setting $I_\lambda := \lambda - 1 - \log\lambda$ ($\lambda > 0,\ \lambda \neq 1$), and writing $\zeta_\lambda \in (0,1)$ for the survival probability of a Galton–Watson process with Poisson(λ) offspring distribution, the main results that we shall (re)establish along the way are summarised in the following table.

Regime	$\boldsymbol{\lambda}$	**Behaviour of $\lvert\mathcal{C}_{\max}\rvert$ as $n \to \infty$**
Subcritical	$\lambda < 1$	$\dfrac{\lvert\mathcal{C}_{\max}\rvert}{\log n} \xrightarrow{\mathbb{P}} \dfrac{1}{I_\lambda}$, with matching upper and lower bounds: for every $a \gtrless 1/I_\lambda$ there exists $\delta > 0$ such that $\mathbb{P}(\lvert\mathcal{C}_{\max}\rvert \gtrless a \log n) = O(n^{-\delta})$.
Critical	$\lambda = 1$	$\lvert\mathcal{C}_{\max}\rvert$ is of order $n^{2/3}$ with high probability, with heavy-tailed fluctuations of the same order.
Supercritical	$\lambda > 1$	$\dfrac{\lvert\mathcal{C}_{\max}\rvert}{n} \xrightarrow{\mathbb{P}} \zeta_\lambda \in (0,1)$, with Gaussian fluctuations of order $\sqrt{n}$; the giant component is unique, and once it is removed the rest of the graph behaves like a subcritical Erdős–Rényi graph.

The thesis is organised as follows. Chapter 2 introduces the basic graph-theoretic background and the Erdős–Rényi model, together with the large-deviation constant I_λ that controls the subcritical regime. Chapter 3 develops the theory of Galton–Watson branching processes needed in the sequel: the extinction-versus-survival

dichotomy, the total progeny and its exponential tail bound, and the binomial and Poisson offspring distributions that arise naturally in the comparison with $G(n, p)$. Chapter 4 contains the core of the thesis. We establish stochastic domination bounds between the breadth-first exploration and suitable branching processes, and use them to analyse the three regimes: the law of large logarithms in the subcritical case, the law of large numbers and uniqueness of the giant component in the supercritical case, and the $n^{2/3}$ scaling in the critical window. Chapter 5 concludes with a few comments on further directions, and two appendices gather the necessary background material.

Chapter 2

Preliminaries

This chapter introduces the graph-theoretic notation used throughout the thesis. Unless otherwise stated, all graphs are finite, simple, and undirected.

The Erdős–Rényi (ER) model, which we shall introduce formally in Section 2.2, is the classical starting point of random graph theory. It provides the simplest probabilistic framework in which fundamental phenomena such as thresholds, phase transitions, and the emergence of a giant component can be studied in a rigorous way. In this work we mainly follow the modern exposition of van der Hofstad [41], which develops these ideas using probabilistic tools closely related to branching processes.

Random graphs were introduced independently by Erdős and Rényi in a series of seminal papers [16, 17] and by Gilbert [20]. Their work revealed that many global properties of graphs undergo sharp transitions as the edge probability varies. Among the most striking of these phenomena is the sudden appearance of a connected component containing a positive fraction of all vertices, the so-called *giant component*.

2.1 Graph Theory Basics

We begin by recalling fundamental graph-theoretic definitions and notation that will be used throughout this work, particularly in the analysis of Erdős–Rényi random graphs. Our presentation follows classical references such as Diestel [12], West [43], and van der Hofstad [41].

Definition 2.1 (Simple graph)**.** A *graph* is an ordered pair $G = (V, E)$ where V is a finite set of *vertices* and E is a collection of unordered pairs $\{u, v\}$ with distinct $u, v \in V$. No multiple edges or loops are allowed, hence the adjective *simple*.

The simplest and most symmetric example is the complete graph.

Definition 2.2 (Complete graph)**.** For $n \geq 1$ we denote by K_n the unique simple graph on the vertex set $[n] = \{1, \ldots, n\}$ whose edge set contains *all* unordered pairs of distinct vertices,

$$E(K_n) = \big\{\{i, j\} \colon 1 \leq i < j \leq n\big\}.$$

Thus K_n has $\binom{n}{2}$ edges and is maximal connected.

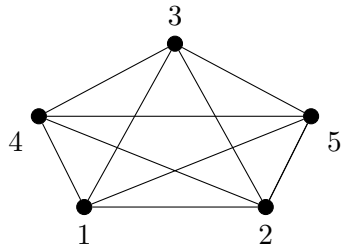


Figure 2.1: The complete graph K_5. Every pair of vertices is adjacent.

Every simple graph on n vertices can be obtained from K_n by deleting edges. In this sense K_n serves as a universal reference.

A basic quantity attached to a vertex is its degree, which records how many neighbors it has. Although the degree is a local notion, it often provides insight into global features of large graphs, such as density, sparsity, or the shape of the degree distribution.

Definition 2.3 (Vertex degree)**.** Let $G = (V, E)$ be a graph. The *degree* of a vertex $u \in V$ is

$$d_u = \#\{\, v \in V : \{u, v\} \in E \,\}.$$

A vertex with $d_u = 0$ is called *isolated*.

Example 2.4 (Vertex degree in a complete graph)**.** Consider the complete graph K_5 in Figure 2.1. Every vertex is adjacent to the other four, hence

$$d_u = 4 \quad \text{for all } u \in V(K_5).$$

This shows that in a complete graph all vertices have the same degree.

We next introduce subgraphs and induced subgraphs, which will be used when studying connected components.

Definition 2.5 (Subgraph and induced subgraph)**.** Let $G = (V, E)$ be a graph. A *subgraph* is a graph $H = (V_H, E_H)$ with $V_H \subseteq V$ and $E_H \subseteq E$. If E_H contains all edges of G with both endpoints in V_H, then $H = G[V_H]$ is called the *induced subgraph* on V_H.

Example 2.6 (Subgraph vs. induced subgraph)**.** Consider the graph G with vertex set $V = \{1, 2, 3, 4\}$ and edge set

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

If we take $V_H = \{1, 2, 3\}$, then:

- choosing $E_H = \{\{1,2\},\{2,3\}\}$ gives a *subgraph*,
- choosing $E_H = \{\{1,2\},\{2,3\},\{3,1\}\}$ gives the *induced subgraph* $G[V_H]$, namely the triangle K_3.

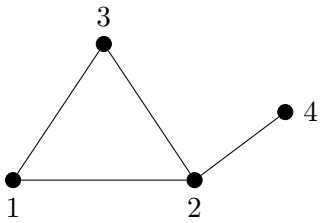


Figure 2.2: Graph G with vertices $\{1,2,3,4\}$. The induced subgraph on $\{1,2,3\}$ is the triangle K_3.

Paths and cycles provide the basic language for connectivity and graph distance.

Definition 2.7 (Path and cycle [12, Sec. 1.3])**.** Let $G = (V, E)$ be a graph.

A *path* in G is a sequence of vertices

$$P = (x_0, x_1, \ldots, x_k),$$

such that $\{x_{i-1}, x_i\} \in E$ for every $i = 1, \ldots, k$. Its *length* is k (the number of edges). We often denote the same path by concatenation,

$$P = x_0 x_1 \cdots x_k.$$

If all vertices $x_0, \ldots, x_k$ are distinct, the path is called *simple*. The vertices x_0 and x_k are called the *endpoints* of P.

A *cycle* is a path $(x_0, x_1, \ldots, x_k)$ with $k \geq 3$ such that $x_0 = x_k$ and $x_0, x_1, \ldots, x_{k-1}$ are all distinct.

Example 2.8 (A path as a subgraph)**.** Let G be the graph shown in Figure 2.3. The bold edges form a path

$$P = (x_0, x_1, \ldots, x_6),$$

of length 6, with endpoints x_0 and x_6. This illustrates that a path is not necessarily the whole graph, but may appear as a subgraph embedded in a larger structure.

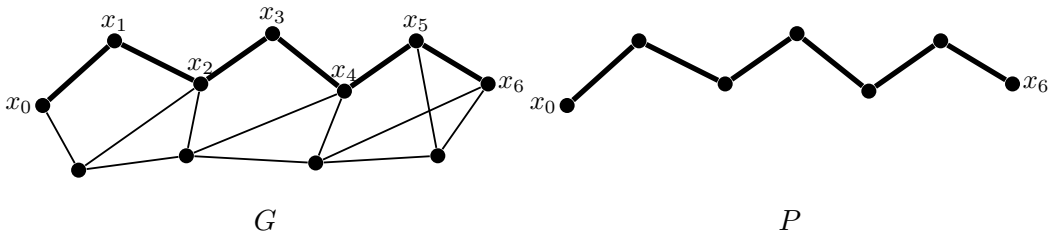


Figure 2.3: A path $P = P^6$ embedded in a graph G, highlighted as a subgraph.

Definition 2.9 (Distance)**.** For $u, v \in V$, the *graph distance* $\mathrm{dist}_G(u, v)$ is the length of a shortest path between u and v, or ∞ if no path exists.

With paths one can define connectivity.

Definition 2.10 (Connectedness and components)**.** A graph is *connected* if any two vertices can be joined by a path. A *maximal* connected subgraph is called a *connected component*.

Closely related is the graph distance, which measures how far apart two vertices are in terms of path length.

Example 2.11 (An illustrative graph)**.** Consider the graph G in Figure 2.4, with vertex set $V = \{A, B, C, D, E\}$ and edge set $E = \{\{A, B\}, \{B, C\}, \{B, D\}, \{D, A\}\}$. This graph has two components: $\{A, B, C, D\}$ and the isolated vertex E. The degrees are $d_A = 2,\ d_B = 3,\ d_C = 1,\ d_D = 2,\ d_E = 0$. Vertices A, B, D form a 3-cycle. Distances include $\mathrm{dist}_G(C, A) = 2$ via the path $C - B - A$ and $\mathrm{dist}_G(C, D) = 2$ via the path $C - B - D$. The isolated vertex E has no path to the others, so $\mathrm{dist}_G(E, A) = \infty$.

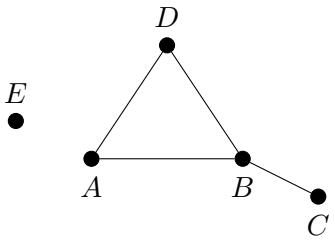


Figure 2.4: A graph with two components: a 3-cycle plus vertex C, and the isolated vertex E.

Finally, we turn to a concept that will play a central role in random graph theory: the existence of a "giant component" in a sequence of graphs. The previous example illustrates how degree, paths, distance, and connected components describe the structure of a single, fixed graph. In random graph theory, however, one is typically interested not in one specific graph but in a *sequence* of graphs whose size grows.

More precisely, many global phenomena only emerge when the number of vertices tends to infinity. In this asymptotic viewpoint, properties of interest are no longer deterministic but probabilistic, and one studies how typical graphs in a given model behave as $n \to \infty$. A central question is how the sizes of connected components scale with the total number of vertices.

This leads to the notion of a *giant component*, which formalises the idea that, in large graphs, a positive fraction of all vertices may belong to a single connected component, while all remaining components are comparatively small.

Definition 2.12 (Highly connected sequence)**.** A sequence of graphs $(G_n)_{n\geq 1}$ with $|V(G_n)| = n$ is called *highly connected* if, as $n \to \infty$,

$$\frac{|\mathbb{C}_{\max}(G_n)|}{n} \longrightarrow c > 0, \qquad \frac{|\mathbb{C}_{(2)}(G_n)|}{n} \longrightarrow 0,$$

where $\mathbb{C}_{\max}$ and $\mathbb{C}_{(2)}$ denote respectively the largest and second–largest components.

When (G_n) is a random graph sequence, the above limits are understood to hold in probability (or, equivalently in our setting, with high probability).

The concept of a *highly connected sequence* formalizes the idea that in many large random graphs almost all vertices belong to one "giant" component while all other components are negligible in size. This phenomenon is one of the central themes of modern random graph theory and will play a key role throughout this work. Understanding when and how such a giant component emerges is the fundamental question that motivates the study of random graphs.

The natural mathematical framework where this phenomenon first appeared is the Erdős–Rényi model. Indeed, Erdős and Rényi themselves discovered in their pioneering papers that $G(n,p)$ undergoes a *phase transition*: as the edge probability p crosses a critical threshold of order $1/n$, the graph abruptly changes from having only small tree-like components to containing one macroscopic component of linear size.

In the next section we therefore introduce the Erdős–Rényi random graph in detail, recalling its historical origins, definitions, and the basic threshold results that make it the cornerstone of the theory.

2.2 Erdős–Rényi Random Graphs

We now turn to the formal definition of the model that will be the central object of study in the rest of this thesis.

Definition 2.13 (Erdős–Rényi random graph [41, Sec. 4.1])**.** Let $n \in \mathbb{N}$ and $p \in [0,1]$. The Erdős–Rényi random graph $G(n,p)$ is the simple graph on the vertex set $[n] := \{1,\dots,n\}$ obtained by connecting each pair of distinct vertices $i, j \in [n]$, $i \neq j$, independently with probability p. Equivalently, every one of the $\binom{n}{2}$ possible edges is included in the graph independently with probability p.

Throughout this thesis we focus on the *sparse regime*

$$p = \frac{\lambda}{n}, \qquad \lambda > 0 \text{ fixed},$$

and study the asymptotic behavior of $G(n,p)$ as $n \to \infty$. This scaling is called *sparse* because the number of edges is much smaller than the maximal possible number $\binom{n}{2}$. Moreover, this scaling is natural, as for any fixed vertex v the degree satisfies

$$\deg(v) \sim \mathrm{Bin}(n-1,p), \qquad \mathbb{E}[\deg(v)] = (n-1)p \longrightarrow \lambda,$$

so that the graph remains locally finite in the limit. Moreover, this is precisely the scale at which a phase transition in the component structure occurs.

We distinguish three regimes:

$$\begin{cases} \textit{subcritical} & \text{if } \lambda < 1, \\ \textit{critical} & \text{if } \lambda = 1, \\ \textit{supercritical} & \text{if } \lambda > 1. \end{cases}$$

The behavior of connected components differs dramatically across these regimes. To describe this behavior rigorously, we first introduce the relevant notions of connected components and giant components.

Definition 2.14 (Connected component)**.** For a graph $G = (V, E)$ and a vertex $v \in V$, the *connected component* of v, denoted by $\mathcal{C}(v)$, is the maximal connected subgraph of G containing v. We write $\mathcal{C}_{\max}$ for the largest connected component of G.

Informally, a *giant component* is a connected component whose size is of linear order in n, that is, containing a positive fraction of all vertices.

In the subcritical regime $\lambda < 1$, all components are typically small and tree–like, with sizes of order $\log n$. In contrast, when $\lambda > 1$, a unique giant component emerges, while all remaining components are much smaller. At the critical point $\lambda = 1$, component sizes exhibit large fluctuations and are of order $n^{2/3}$.

The comparison between component exploration and Galton–Watson processes is developed in Chapters 3 and 4.

2.2.1 A large-deviation constant for Poisson variables

A quantity that will play an important role in the subcritical regime of Erdős–Rényi random graphs is the constant

$$I_\lambda \;=\; \lambda - 1 - \log \lambda, \qquad \lambda > 0.$$

It satisfies $I_\lambda > 0$ for all $\lambda < 1$ and vanishes at $\lambda = 1$.

This constant arises naturally from large deviation theory for Poisson random variables. More precisely, let $X \sim \mathrm{Pois}(\lambda)$. The moment generating function of X is

$$\mathbb{E}[e^{\theta X}] = \exp\big(\lambda(e^\theta - 1)\big),$$

so that the logarithmic moment generating function is

$$\Lambda(\theta) = \log \mathbb{E}[e^{\theta X}] = \lambda(e^\theta - 1).$$

Large deviation theory (via the Chernoff bound) shows that probabilities of the form $\mathbb{P}(X \geq x)$ decay exponentially with rate given by the Legendre transform of Λ, namely

$$I(x) = \sup_{\theta \geq 0} \big(\theta x - \Lambda(\theta)\big),$$

see Durrett [15]. Evaluating this rate function at $x = 1$ yields

$$I(1) = \sup_{\theta \geq 0} \big(\theta - \lambda(e^{\theta} - 1)\big).$$

Optimising over θ gives $e^{\theta} = 1/\lambda$, and hence

$$I(1) = -\log \lambda - (1 - \lambda) = \lambda - 1 - \log \lambda.$$

Thus the constant I_λ is precisely the large-deviation rate function of a Poisson(λ) random variable evaluated at 1. This quantity appears naturally in exponential tail bounds for component sizes in subcritical Erdős–Rényi graphs, see Durrett [15, Sec. 1.3]. It will later determine the exponential decay of the total progeny of the associated branching process, and hence the typical size of connected components in the subcritical regime.

Chapter 3

Galton–Watson Branching Processes

The origins of the Galton–Watson process date back to the late nineteenth century, when Francis Galton proposed a probabilistic model to describe how family surnames might disappear over generations due to random variation in the number of male descendants [42]. Several decades later, in the 1930s, Leo Szilard independently rediscovered the same mathematical structure while studying the behaviour of neutrons in a nuclear fission chain reaction. The probabilistic mechanism determining whether a family line eventually dies out or survives indefinitely turned out to be mathematically equivalent to the question of whether a nuclear reaction would extinguish or become self–sustaining. This realization established the Galton–Watson process as a fundamental model for understanding growth and extinction phenomena in random systems. Since then, it has become one of the cornerstones of stochastic modelling, with applications ranging from genetics and epidemiology to physics and network theory [see, e.g., 5, 22, 23, 41].

In its simplest form, the model describes the evolution of a population observed at discrete time steps $n = 0, 1, 2, \ldots$. At time $n = 0$, the process starts from an initial population size Z_0, often taken to be $Z_0 = 1$, corresponding to a single ancestor. Each individual in generation n then produces independently a random number of offspring in generation $n + 1$, possibly zero, according to a common probability distribution. If X denotes the offspring random variable and

$$p_k = \mathbb{P}(X = k),$$

then p_k represents the probability that one individual has exactly k children. The number of individuals in generation n is denoted by the random variable Z_n. Starting from the initial condition Z_0, the sequence $(Z_n)_{n\geq 0}$ evolves recursively, since every individual in generation n independently gives rise to a random number of descendants in generation $n + 1$.

This recursive structure makes the Galton–Watson process a natural probabilistic framework for describing self–replicating systems, whether they correspond to

biological populations, particles in physical reactions, or the local growth of clusters in random graphs [21, 27].

From a mathematical perspective, branching processes form one of the simplest stochastic models of population dynamics. They provide a natural way to study how a family tree evolves when each individual reproduces independently according to a fixed offspring distribution. This simplicity makes them not only analytically tractable but also extremely powerful for capturing the qualitative behavior of systems driven by random reproduction or growth [5].

The motivation for introducing branching processes in this work lies in their close connection with Erdős–Rényi random graphs. Exploring the connected component of a vertex in $G(n,p)$ can be coupled with a Galton–Watson process: the number of "children" corresponds to the number of new neighbours discovered at each exploration step [41, Ch. 3–4]. Consequently, the extinction or survival of the branching process mirrors the absence or presence of a giant component in the random graph. This coupling provides the key to understanding the phase transition in $G(n,p)$ and justifies the use of branching processes as a fundamental analytical tool in our study.

3.1 Definitions and first properties

Definition 3.1 (Galton–Watson process)**.** Let X be a non-negative integer-valued random variable, called the *offspring random variable*, with mean $\mu = \mathbb{E}[X] \in [0,\infty]$. A Galton–Watson process is the sequence $(Z_t)_{t\geq 0}$ defined by

$$Z_0 = 1, \qquad Z_t = \sum_{i=1}^{Z_{t-1}} X_{t,i}, \quad t \in \mathbb{Z}^+,$$

where the time index t ranges over the non-negative integers (so the process is indexed by discrete time), and where the random variables $(X_{t,i})_{t,i}$ are independent and identically distributed copies of X. We write $\mathrm{GW}(X)$ for its law.

The random variable Z_t denotes the number of individuals in the t-th generation of the population. Intuitively, the process starts with a single ancestor at time $t = 0$, and each individual in generation t gives birth independently to a random number of offspring distributed as X.

The collection of probabilities

$$p_k = \mathbb{P}(X = k), \qquad k = 0, 1, 2, \ldots,$$

is called the *offspring distribution*. It specifies the probability that a single individual produces exactly k children. Throughout this chapter, we will often identify the offspring distribution with the law of the random variable X.

The process (Z_t) is in fact a *Markov chain* (see Appendix A.1): given the present generation, the distribution of the next generation is independent of the past [5, Sec. 1.1]. From this follows the *branching property*: conditional on $Z_1 = i$, the

descendants of these i individuals evolve independently, each according to the same law as the original process. The branching property allows many quantities of interest, such as extinction probabilities and total progeny, to be analysed recursively through conditioning arguments and generating functions. It will also play a central role later when comparing branching processes with the exploration of connected components in Erdős–Rényi random graphs.

A convenient way to encode the offspring distribution is through its *probability generating function*. Besides compactly storing the probabilities $p_k = \mathbb{P}(X = k)$, the probability generating function allows one to compute moments and plays a central role in the analysis of extinction. For a general discussion of generating functions in branching processes, see van der Hofstad [41, Sec. 3.1] and Harris [22, Ch. 1].

Remark 3.2 (Basic properties of the probability generating function)**.** Let X be a non-negative integer-valued random variable with $p_k = \mathbb{P}(X = k)$ and *probability generating function* (PGF) $G_X(s) = \mathbb{E}[s^X] = \sum_{k\geq 0} p_k s^k$ for $s \in [0, 1]$. Then:

- $G_X(0) = \mathbb{P}(X = 0)$ and $G_X(1) = 1$.
- G_X is non-decreasing and convex on $[0, 1]$.
- If $\mathbb{E}[X] < \infty$, then

$$G'_X(s) = \mathbb{E}[Xs^{X-1}] \quad \text{for } s \in (0, 1), \qquad \text{hence} \qquad G'_X(1) = \mathbb{E}[X].$$

The proofs of these properties are standard and can be found, e.g., in van der Hofstad [41, Sec. 3.1] and Athreya and Ney [5, Ch. II].

We now formalize the notion of extinction for a Galton–Watson process. Since $(Z_n)_{n\geq 0}$ is absorbed at 0, extinction occurs if the population eventually reaches 0, i.e.

$$\{\text{extinction}\} = \{\exists\, n \geq 0 : Z_n = 0\}.$$

The corresponding extinction probability is defined by

$$\eta := \mathbb{P}(\exists\, n \geq 0 : Z_n = 0). \tag{3.1}$$

Since (Z_n) is absorbed at 0, this event means that the population eventually dies out. The key result (which we will prove in Theorem 3.3) is that η is characterized as the smallest fixed point of G_X [see also 5, Ch. I.5].

Intuitively, the mean offspring number $\mu = \mathbb{E}[X]$ should govern the long-term behavior: if $\mu < 1$, each generation is smaller on average and extinction is expected, whereas if $\mu > 1$, there is enough "reproductive capacity" to allow survival with positive probability. We therefore introduce the following terminology: a Galton–Watson process is called *subcritical* if $\mu < 1$, *critical* if $\mu = 1$, and *supercritical* if $\mu > 1$. The names reflect precisely this dichotomy: in the subcritical case, reproduction is insufficient to sustain the population; in the supercritical case, reproduction exceeds

the replacement threshold; and the critical case $\mu = 1$ lies exactly at the boundary between these two regimes. A convenient tool to analyze this dichotomy is the PGF of the offspring distribution, which turns the extinction problem into a fixed-point equation for G_X [see, e.g., 41, Sec. 3.1].

Theorem 3.3 (Survival vs. extinction)**.** *Let $(Z_n)_{n\geq 0}$ be a Galton–Watson branching process with i.i.d. offspring distribution X and offspring PGF*

$$G_X(s) = \mathbb{E}[s^X].$$

Then the quantity η defined in (3.1) *is the smallest solution in $[0, 1]$ of the fixed-point equation*

$$\eta = G_X(\eta).$$

Moreover,

$$\eta = \begin{cases} 0, & \text{if } \mathbb{P}(X = 1) = 1, \\ 1, & \text{if } \mathbb{E}[X] < 1, \\ 1, & \text{if } \mathbb{E}[X] = 1 \text{ and } \mathbb{P}(X = 1) < 1, \\ < 1, & \text{if } \mathbb{E}[X] > 1. \end{cases}$$

Proof. The classical argument, which we follow here, is due to Steffensen and is developed in detail in Athreya and Ney [5, Ch. I] and van der Hofstad [41, Ch. 3]. The proof is structured in three steps. In Step 1 we derive a recursion for the generating functions $G_n(s) = \mathbb{E}[s^{Z_n}]$ and, specializing it at $s = 0$, an iterative equation for the finite–time extinction probabilities $\eta_n = \mathbb{P}(Z_n = 0)$. In Step 2 we pass to the limit and show that η is a fixed point of G_X, and prove by induction that it is the *smallest* such fixed point in $[0, 1]$. Finally, in Step 3 we analyze the number of fixed points of G_X through convexity arguments, and obtain the trichotomy in terms of $\mathbb{E}[X]$.

For each $n \geq 0$ set

$$\eta_n := \mathbb{P}(Z_n = 0).$$

Since the event $\{Z_n = 0\}$ implies $\{Z_{n+1} = 0\}$, the sequence (η_n) is non–decreasing and converges to

$$\eta = \lim_{n\to\infty} \eta_n = \mathbb{P}(Z_n = 0 \text{ for some } n),$$

which is precisely the extinction probability. Let

$$G_n(s) := \mathbb{E}[s^{Z_n}], \qquad s \in [0, 1],$$

be the PGF of Z_n. In particular,

$$\eta_n = G_n(0), \tag{3.2}$$

since for an integer–valued random variable Y we have $\mathbb{P}(Y = 0) = \mathbb{E}[0^Y] = G_Y(0)$.

Step 1: Recursion for the PGFs. Condition on the first generation. Writing $p_i = \mathbb{P}(X = i)$ and using $\mathbb{P}(Z_1 = i) = p_i$, we obtain, for $n \geq 2$, $G_n(s) = \mathbb{E}[s^{Z_n}] =$

$\sum_{i=0}^{\infty} p_i \, \mathbb{E}\big[s^{Z_n} \mid Z_1 = i\big]$. Given $Z_1 = i$, the i individuals in generation 1 reproduce independently, each giving rise to a copy of the process started from one individual; hence Z_n is the sum of i i.i.d. copies of Z_{n-1} and $\mathbb{E}\big[s^{Z_n} \mid Z_1 = i\big] = \big(G_{n-1}(s)\big)^i$. Therefore

$$G_n(s) = \sum_{i=0}^{\infty} p_i \big(G_{n-1}(s)\big)^i = G_X\big(G_{n-1}(s)\big), \tag{3.3}$$

where $G_X(s) = \sum_{i \geq 0} p_i s^i$ is the probability generating function of X. Taking $s = 0$ in (3.3) and using (3.2) gives, for $n \geq 1$,

$$\eta_n = G_X(\eta_{n-1}), \qquad \eta_0 = \mathbb{P}(Z_0 = 0) = 0. \tag{3.4}$$

Step 2: Fixed–point equation and minimality. By continuity of G_X on $[0,1]$ and the monotone convergence of (η_n), passing to the limit in (3.4) yields $\eta = G_X(\eta)$, so η is a fixed point of G_X. We now prove that η is in fact the *smallest* fixed point of G_X in $[0,1]$. Let $\psi \in [0,1]$ be any solution of $\psi = G_X(\psi)$. We claim that $\eta \leq \psi$, and we prove $\eta_n \leq \psi$ for all n by induction. For $n = 0$ we have $\eta_0 = 0 \leq \psi$. Assume $\eta_{n-1} \leq \psi$ for some $n \geq 1$. Since X is non-negative, G_X is non-decreasing on $[0,1]$ (cf. Remark 3.2), hence $\eta_n = G_X(\eta_{n-1}) \leq G_X(\psi) = \psi$. We infer that $\eta_n \leq \psi$ for all n, thus, taking limits gives $\eta \leq \psi$. Therefore η is the *smallest* fixed point of G_X in $[0,1]$.

Step 3: Number of fixed points and the role of $\mathbb{E}[X]$. We next determine when this smallest fixed point equals 1 and when it is < 1.

First consider the trivial cases:

- If $\mathbb{P}(X = 1) = 1$, then $Z_n = 1$ for all n, so extinction never occurs and $\eta = 0$. This degenerate case is usually excluded.

- If $\mathbb{P}(X \leq 1) = 1$ and $\mathbb{P}(X = 0) > 0$, then each generation consists of at most one individual; the single line of descent is killed with probability p at each step, so
$$\mathbb{P}(Z_n = 0) = 1 - (1-p)^n,$$
which tends to 1 as $n \to \infty$ whenever $p > 0$. Hence $\eta = 1$.

In the remainder we assume $\mathbb{P}(X \leq 1) < 1$, which is equivalent to $\mathbb{P}(X \geq 2) > 0$. This implies that G_X is *strictly* convex on $(0,1)$, since

$$G_X''(s) = \mathbb{E}\big[X(X-1)s^{X-2}\big] > 0, \qquad 0 < s < 1,$$

the expectation being strictly positive precisely because $\mathbb{P}(X \geq 2) > 0$.

Since $G_X(1) = 1$, the equation $s = G_X(s)$ always has $s = 1$ as a solution, and strict convexity implies that there can be at most two solutions in $[0,1]$, since $G_X(1) = 1$ and $G_X(0) = \mathbb{P}(X = 0) \geq 0$. Moreover, the derivative at 1 is $G_X'(1) = \mathbb{E}[X]$, and we note that this is precisely the mean number of offspring, which will drive the three cases below.

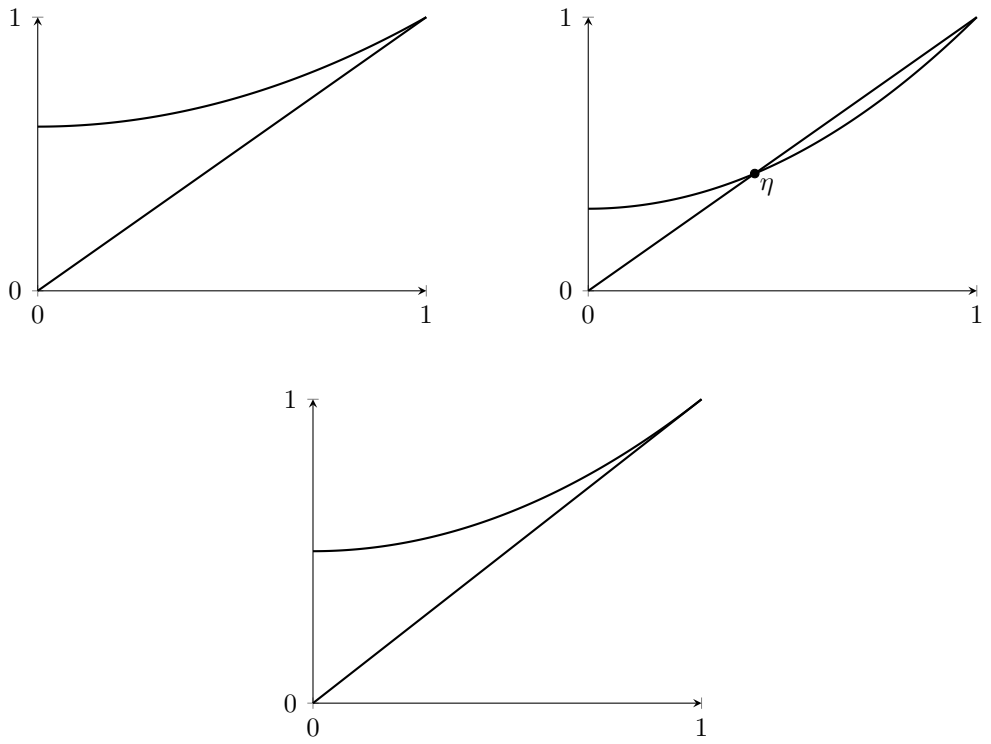


Figure 3.1: Qualitative shapes of the solutions to $s = G_X(s)$ in the cases $\mathbb{E}[X] < 1$ (top left, subcritical), $\mathbb{E}[X] > 1$ (top right, supercritical), and $\mathbb{E}[X] = 1$ (bottom, critical).

- If $\mathbb{E}[X] < 1$, then $G'_X(1) < 1$, so the graph of G_X crosses the diagonal $s \mapsto s$ from above at $s = 1$, and there is no other intersection in $[0, 1]$. Thus $s = 1$ is the unique fixed point, and by minimality we must have $\eta = 1$.

- If $\mathbb{E}[X] > 1$, then $G'_X(1) > 1$, and strict convexity forces the graph of G_X to intersect the diagonal at some point $\eta \in (0, 1)$ and again at 1. Hence there are two fixed points, one in $(0, 1)$ and one at 1; the smallest fixed point is $\eta < 1$, so the extinction probability is strictly less than 1.

- If $\mathbb{E}[X] = 1$ and $\mathbb{P}(X = 1) < 1$, then again convexity implies that $s = 1$ is the unique solution of $s = G_X(s)$ in $[0, 1]$. Hence $\eta = 1$.

Combining these cases yields the stated trichotomy for η in terms of $\mathbb{E}[X]$; Figure 3.1 illustrates the three cases geometrically. ■

Remark 3.4. Note that $G_X(1) = 1$ always, so $s = 1$ is always a fixed point of G_X. In the supercritical case $\mathbb{E}[X] > 1$, the fixed–point equation $s = G_X(s)$ has an additional solution $\eta \in (0, 1)$, as illustrated in the top–right panel of Figure 3.1.

It is often convenient to work with the *survival probability*

$$\zeta := \mathbb{P}(Z_n > 0 \text{ for all } n \geq 0).$$

Since extinction and survival are complementary events for a Galton–Watson process (excluding the degenerate case $X \equiv 1$ a.s., where the process never becomes extinct but also never "survives" in a non–trivial sense), we have the identity

$$\zeta \;=\; 1 - \eta. \tag{3.5}$$

By Theorem 3.3, we therefore have $\zeta > 0$ if and only if $\mathbb{E}[X] > 1$ (excluding the degenerate case $X \equiv 1$), while $\zeta = 0$ in the subcritical and (non-degenerate) critical regimes [5, Ch. I.5].

A key example for random graphs is the Poisson offspring distribution $X \sim \mathrm{Pois}(\lambda)$. Here the offspring PGF is $G_X(s) = \exp(\lambda(s-1))$, and the extinction probability is the smallest solution of

$$\eta \;=\; e^{\lambda(\eta-1)}.$$

Consequently, for $\lambda \leq 1$ we have $\zeta_\lambda = 0$, whereas for $\lambda > 1$ the survival probability $\zeta_\lambda = 1 - \eta$ is strictly positive and increases to 1 as $\lambda \to \infty$. Figure 3.2 illustrates the shape of ζ_λ as a function of λ. This equation also arises in classical mean-field percolation and in the analysis of epidemic models [14].

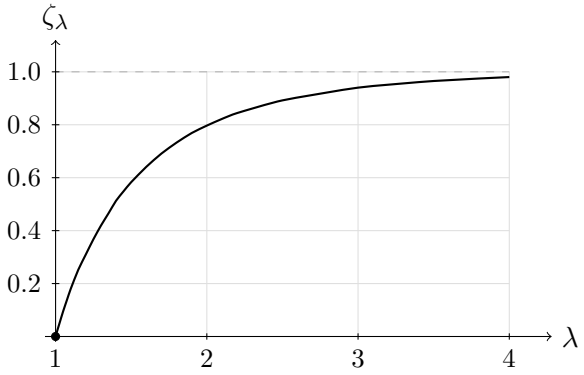


Figure 3.2: Survival probability ζ_λ of a $\mathrm{Pois}(\lambda)$ Galton–Watson process as a function of $\lambda \geq 1$. The dashed line marks the asymptote $\zeta_\lambda \to 1$ as $\lambda \to \infty$.

3.2 Total progeny

We now study the total progeny, which will be compared with the size of a connected component in Chapter 4.

We thus study the *total progeny*

$$T \;=\; \sum_{n=0}^{\infty} Z_n,$$

i.e. the total number of individuals ever born in the branching process, and we denote by $G_T(s) = \mathbb{E}[s^T]$ its probability generating function, defined for $s \in [0, 1]$.

Theorem 3.5 (Total-progeny PGF)**.** *For a Galton–Watson process with offspring PGF G_X, the function G_T satisfies, for all $s \in [0, 1]$,*

$$G_T(s) \;=\; s\, G_X(G_T(s)).$$

Proof. We condition on the size $Z_1 = i$ of the first generation. If $Z_1 = i$, then for each $j = 1, \dots, i$, let T_j denote the total progeny of the j-th child of the root.

By the *branching property*, conditionally on $Z_1 = i$, the random variables $T_1, T_2, \dots, T_i$ are independent and identically distributed, each having the same distribution as T. Therefore, conditioned on $Z_1 = i$, the total progeny satisfies

$$T \;=\; 1 \,+\, \sum_{j=1}^{Z_1} T_j, \tag{3.6}$$

where the empty sum for $Z_1 = 0$ is taken to be 0 [41, Eq. (3.1.20)]. Using the definition of the probability generating function,

$$G_T(s) \;=\; \mathbb{E}[s^{\,T}] \;=\; \sum_{i=0}^{\infty} p_i\, \mathbb{E}[s^{\,T} \mid Z_1 = i].$$

Conditioning on $Z_1 = i$ and applying (3.6), we obtain

$$\begin{aligned}\mathbb{E}[s^{\,T} \mid Z_1 = i] &= \mathbb{E}\left[s^{\,1+\sum_{j=1}^{i} T_j}\right] = s\,\mathbb{E}\left[s^{\sum_{j=1}^{i} T_j}\right] \\ &= s\,\left(\mathbb{E}[s^{\,T}]\right)^i = s\,\left(G_T(s)\right)^i,\end{aligned}$$

since $T_1, \dots, T_i$ are i.i.d. Substituting into the previous expression for $G_T(s)$, we get

$$G_T(s) = s \sum_{i=0}^{\infty} p_i\, \left(G_T(s)\right)^i = s\, G_X(G_T(s)),$$

which is the desired functional equation. ∎

3.2.1 Expected generation size

We end this chapter by computing the expected generation sizes, as it will be pivotal in the analysis of the Erdős–Rényi random graph and to explain the emergence of the giant component via the survival of the branching process.

The independence and identical distribution of the random variables $X_{t,i}$ imply that all individuals reproduce according to the same probabilistic rule, independently of the population size in the previous generation. This property allows us to derive

simple recursive relations for the expected number of individuals in each generation [5, Ch. I.3]. In particular, we can express the size of generation n in terms of the previous one as

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_{n,i},$$

where Z_{n-1} is the number of individuals in generation $n-1$ and $X_{n,i}$ denotes the number of children produced by the i-th individual of that generation. The variables $X_{n,i}$ are i.i.d. copies of the offspring variable X and are independent of Z_{n-1}. This recursive representation is central to the analysis of the process and will now be used to compute its expectation.

Proposition 3.6 (Expectation of generation sizes)**.** *Let $(Z_n)_{n\geq 0}$ be a Galton–Watson process with offspring distribution X and mean $\mu = \mathbb{E}[X] \in [0,\infty)$. Then, for all $n \geq 0$,*

$$\mathbb{E}[Z_n] \;=\; \mu^{\,n}. \tag{3.7}$$

Proof. Let $(Z_n)_{n\geq 0}$ be a Galton–Watson process with offspring variable X and mean $\mu = \mathbb{E}[X]$. Recall the basic decomposition

$$Z_n \;=\; \sum_{i=1}^{Z_{n-1}} X_{n,i}, \qquad n \geq 1, \tag{3.8}$$

where $\{X_{n,i} : n \geq 1, i \geq 1\}$ are i.i.d. copies of X and are independent of Z_{n-1}.

Fix $n \geq 1$. For any $m \in \mathbb{N}_0$, conditioning on $\{Z_{n-1} = m\}$ and using linearity of expectation and the i.i.d. property of the $X_{n,i}$,

$$\mathbb{E}[Z_n \mid Z_{n-1} = m] = \sum_{i=1}^{m} \mathbb{E}[X_{n,i} \mid Z_{n-1} = m] = m\,\mu.$$

Therefore,

$$\mathbb{E}[Z_n \mid Z_{n-1}] \;=\; \mu\, Z_{n-1} \qquad \text{a.s.} \tag{3.9}$$

Taking expectations in (3.9) and using the tower property,

$$\mathbb{E}[Z_n] = \mathbb{E}\big[\,\mathbb{E}[Z_n \mid Z_{n-1}]\,\big] = \mu\,\mathbb{E}[Z_{n-1}].$$

With the initial condition $\mathbb{E}[Z_0] = 1$, this yields by induction

$$\mathbb{E}[Z_n] \;=\; \mu^n, \qquad n \geq 0.$$

■

The previous results provide a complete qualitative picture of the Galton–Watson process. The key parameter is the mean number of offspring $\mu = \mathbb{E}[X]$. The extinction theorem (Theorem 3.3) shows that:

- if $\mu < 1$ (subcritical case), extinction occurs almost surely;

- if $\mu = 1$ and $\mathbb{P}(X = 1) < 1$ (critical case), extinction still occurs almost surely, but with heavy fluctuations;
- if $\mu > 1$ (supercritical case), the population survives with positive probability $1 - \eta$.

At the same time, Theorem 3.6 shows that the expected population size satisfies

$$\mathbb{E}[Z_n] = \mu^n,$$

so that the process grows (or decays) geometrically in expectation, in agreement with the three regimes above.

The total progeny $T = \sum_{n\geq 0} Z_n$ gives a global measure of the size of the tree generated by the process. The functional equation

$$G_T(s) = s\, G_X(G_T(s)),$$

relating the PGF of T to that of the offspring PGF G_X, will be a crucial tool later on: it allows us to control the tail behaviour of T via analytic properties of G_X. In particular, for Poisson offspring $X \sim \text{Pois}(\lambda)$, the large deviation rate function associated with T will determine the logarithmic order of the largest component in the subcritical Erdős–Rényi graph [41, Ch. 4].

From now on, we shall mainly be interested in the case where X has Poisson or binomial distribution with mean $\lambda = \mathbb{E}[X]$. In this setting, a Galton–Watson process provides a natural approximation to the breadth–first exploration of a connected component in the random graph $G(n, p)$ with $p = \lambda/n$. The extinction probability η corresponds to the absence of a giant component, whereas the survival probability $1 - \eta$ will turn out to be the asymptotic fraction of vertices contained in the giant component in the supercritical regime. The next chapter makes this connection precise and uses the branching-process viewpoint developed here to describe the phase transition in Erdős–Rényi random graphs.

3.2.2 The critical case

The critical regime $\mu = \mathbb{E}[X] = 1$ plays a special role. By Theorem 3.3, if $\mathbb{E}[X] = 1$ and $\mathbb{P}(X = 1) < 1$, then the Galton–Watson process becomes extinct almost surely. However, this does not mean that the process is typically small: in the critical case, the total progeny exhibits much larger fluctuations than in the subcritical regime [28, 44].

Let

$$T = \sum_{n\geq 0} Z_n$$

denote the total progeny. Since extinction occurs almost surely, $T < \infty$ almost surely, but its expectation is infinite in the non-degenerate critical case. Indeed, by Proposition 3.6,

$$\mathbb{E}[T] = \sum_{n\geq 0} \mathbb{E}[Z_n] = \sum_{n\geq 0} \mu^n.$$

When $\mu = 1$, this becomes

$$\mathbb{E}[T] = \sum_{n \geq 0} 1 = \infty.$$

Thus, a critical Galton–Watson process dies out almost surely, but it may survive for a long time and produce a large total progeny with non-negligible probability. Kolmogorov's classical asymptotic states that, under a finite variance assumption, $n\,\mathbb{P}(Z_n > 0) \to 2/\sigma^2$ as $n \to \infty$, where $\sigma^2 = \mathrm{Var}(X)$; see Kolmogorov [28], and also Yaglom [44] for the limit distribution of Z_n conditioned on non-extinction. This heavy-tailed behaviour is one of the reasons why the branching-process approximation becomes much more delicate in the critical Erdős–Rényi regime; see, for instance, Harris [22], Athreya and Ney [5], and van der Hofstad [41, Sec. 3.1].

3.2.3 Tail of the total progeny in the subcritical case

We now establish the exponential tail estimate for the total progeny of a subcritical Galton–Watson process. This result will be the key ingredient in the analysis of the largest connected component in the subcritical Erdős–Rényi graph; see Section 4.3.

Proposition 3.7 (Exponential tail of the total progeny)**.** *Let $(Z_n)_{n \geq 0}$ be a Galton–Watson process with offspring distribution $X \sim \mathrm{Pois}(\lambda)$, where $\lambda < 1$. Let $T = \sum_{n \geq 0} Z_n$ denote the total progeny. Then there exists a constant $I_\lambda > 0$, given explicitly by*

$$I_\lambda \;:=\; \lambda - 1 - \log \lambda,$$

such that

$$\mathbb{P}(T \geq k) \;\leq\; e^{-I_\lambda k} \qquad \text{for all } k \geq 1. \tag{3.10}$$

In particular, $\mathbb{E}[T] = \frac{1}{1-\lambda} < \infty$.

Proof. We want to show that $\mathbb{P}(T \geq k) \leq e^{-I_\lambda k}$. The idea is to apply Markov's inequality to an exponential function of T, and then choose the parameter optimally, which is the standard Chernoff–Cramér approach of large deviation theory [11].

For any $\theta > 0$, since $e^{\theta T} \geq e^{\theta k}$ whenever $T \geq k$, Markov's inequality gives

$$\mathbb{P}(T \geq k) \;\leq\; e^{-\theta k}\,\mathbb{E}[e^{\theta T}].$$

To make this bound useful, we need to compute $\mathbb{E}[e^{\theta T}]$. Recall from Theorem 3.5 that the PGF of T satisfies $G_T(s) = s\,G_X(G_T(s))$. Since $\mathbb{E}[e^{\theta T}] = G_T(e^\theta)$, setting $m = G_T(e^\theta)$ and using $G_X(s) = e^{\lambda(s-1)}$ for $X \sim \mathrm{Pois}(\lambda)$, we see that m must satisfy the equation

$$m \;=\; e^\theta\, e^{\lambda(m-1)} \;=\; e^{\theta - \lambda + \lambda m}.$$

When $\lambda < 1$ and $\theta > 0$ is small enough, one can verify that this equation has a finite positive solution $m = m(\theta)$, so $\mathbb{E}[e^{\theta T}] < \infty$. The precise range of θ for which this holds turns out to be $\theta < I_\lambda$, where $I_\lambda = \lambda - 1 - \log\lambda$.

We therefore have, for any $\theta \in (0, I_\lambda)$,

$$\mathbb{P}(T \geq k) \;\leq\; e^{-\theta k}\, m(\theta).$$

Minimising the right-hand side over $\theta \in (0, I_\lambda)$ and using the standard Cramér bound from large deviation theory [11, 41], the infimum equals $e^{-I_\lambda k}$, which gives the desired estimate

$$\mathbb{P}(T \geq k) \;\leq\; e^{-I_\lambda k}.$$

To compute the expected total progeny, we use Theorem 3.6, which gives $\mathbb{E}[Z_n] = \lambda^n$. Since $\lambda < 1$,

$$\mathbb{E}[T] \;=\; \sum_{n=0}^{\infty} \mathbb{E}[Z_n] \;=\; \sum_{n=0}^{\infty} \lambda^n \;=\; \frac{1}{1-\lambda} \;<\; \infty,$$

which is consistent with the exponential tail we just proved. ∎

Remark 3.8. The constant $I_\lambda = \lambda - 1 - \log \lambda$ is the large-deviation rate function of a Poisson(λ) random variable evaluated at 1, i.e. $I_\lambda = I(1)$ where

$$I(x) \;=\; \sup_{s \geq 0} \bigl(sx - \lambda(e^s - 1)\bigr).$$

It satisfies $I_\lambda > 0$ for all $\lambda \neq 1$ and $I_\lambda = 0$ only at $\lambda = 1$; see Dembo and Zeitouni [11, Ch. 2].

3.3 Poisson and binomial branching processes

Throughout this chapter the offspring distribution X has been kept entirely general. In view of the application to the Erdős–Rényi random graph $G(n,p)$ developed in the next chapter, we now focus on the two specific offspring distributions that arise naturally in the analysis: the *binomial* $X \sim \mathrm{Bin}(n,p)$ and the *Poisson* $X \sim \mathrm{Pois}(\lambda)$. When exploring the connected component of a vertex in $G(n,p)$ by a breadth–first search, the number of new neighbours discovered at each step is *roughly* binomial: each currently neutral vertex, i.e. a vertex not yet discovered by the exploration, is connected to the vertex being explored independently with probability p, so that, conditionally on the past, this number is distributed as $\mathrm{Bin}(N,p)$, where N is the number of currently neutral vertices. In the early stages of the exploration N is still close to n, so the number of new neighbours is approximately $\mathrm{Bin}(n,p)$. This is why the exploration can be compared with a binomial Galton–Watson process, as we formalise in Section 4.2. Moreover, in the sparse regime $p = \lambda/n$, the law $\mathrm{Bin}(n, \lambda/n)$ converges to $\mathrm{Pois}(\lambda)$, which is why the Poisson process provides the correct limiting object [41, Ch. 2–3].

3.3.1 Binomial branching processes

Remark 3.9 (PGF and moments of $\mathrm{Bin}(n,p)$)**.** Let $X \sim \mathrm{Bin}(n,p)$. Writing $X = \sum_{i=1}^{n} B_i$ with $B_i \overset{\text{iid}}{\sim} \mathrm{Ber}(p)$ and using the independence of the B_i, we obtain

$$G_X(s) \;=\; \prod_{i=1}^{n} \mathbb{E}[s^{B_i}] \;=\; (1 - p + ps)^n, \qquad s \in [0,1].$$

Moreover, $\mathbb{E}[X] = np$ and $\mathrm{Var}(X) = np(1-p)$.

By Theorem 3.3, the extinction probability η of a GW(Bin(n,p)) process is the smallest solution in $[0,1]$ of

$$\eta \;=\; (1-p+p\,\eta)^n, \tag{3.11}$$

and the process is subcritical, critical, or supercritical according to whether $np<1$, $np=1$, or $np>1$. In the sparse regime $p=\lambda/n$, the equation (3.11) is most conveniently studied through its Poisson limit; see Theorem 3.13 below.

3.3.2 Poisson branching processes

Remark 3.10 (PGF and moments of Pois(λ))**.** Let $X\sim\mathrm{Pois}(\lambda)$. Then

$$G_X(s) \;=\; \sum_{k\ge 0} e^{-\lambda}\frac{\lambda^k}{k!}s^k \;=\; e^{\lambda(s-1)}, \qquad s\in[0,1],$$

and $\mathbb{E}[X]=\mathrm{Var}(X)=\lambda$.

The Poisson PGF leads to a particularly clean fixed-point equation.

Proposition 3.11 (Extinction and survival for Poisson offspring)**.** *Let $(Z_t)_{t\ge0}$ be a* GW(Pois(λ)) *process with extinction probability η_λ. Then η_λ is the smallest solution in $[0,1]$ of*

$$\eta \;=\; e^{\lambda(\eta-1)}. \tag{3.12}$$

Equivalently, the survival probability $\zeta_\lambda := 1-\eta_\lambda$ satisfies

$$\zeta_\lambda \;=\; 1-e^{-\lambda\zeta_\lambda}. \tag{3.13}$$

Moreover:

- *$\zeta_\lambda = 0$ for all $\lambda\le 1$;*
- *$\zeta_\lambda>0$ for all $\lambda>1$, with $\lambda\mapsto\zeta_\lambda$ continuous, strictly increasing on $(1,\infty)$, and $\zeta_\lambda\to 1$ as $\lambda\to\infty$.*

Proof. Equation (3.12) is Theorem 3.3 specialised to $G_X(s)=e^{\lambda(s-1)}$. Substituting $\eta_\lambda=1-\zeta_\lambda$ into (3.12) yields (3.13). The trichotomy for ζ_λ follows directly from Theorem 3.3 since the Pois(λ) process is supercritical if and only if $\lambda>1$. Continuity and strict monotonicity of ζ_λ on $(1,\infty)$ follow from the implicit function theorem applied to $F(\zeta,\lambda):=\zeta-(1-e^{-\lambda\zeta})$. At any (ζ_λ,λ) with $\lambda>1$ and $\zeta_\lambda>0$, the partial derivative satisfies $\partial_\zeta F = 1-\lambda e^{-\lambda\zeta_\lambda}\neq 0$, so ζ_λ is a smooth function of λ. Finally, $\eta_\lambda = e^{\lambda(\eta_\lambda-1)}\le e^{-\lambda\zeta_\lambda}\to 0$ as $\lambda\to\infty$, hence $\zeta_\lambda\to 1$. ∎

Remark 3.12. Equation (3.13) is precisely the equation whose solution ζ_λ gives the fraction of vertices in the giant component of $G(n,\lambda/n)$ in the supercritical regime; see Figure 3.2 for an illustration and Section 4.4 for the precise statement. The same equation appears, with exactly the same probabilistic interpretation, in the mean-field analysis of percolation on infinite trees [see 34, Ch. 5].

3.3.3 From binomial to Poisson: the sparse limit

Theorem 3.13 (Poisson limit theorem)**.** *Let $\lambda > 0$ and $p_n = \lambda/n$. Then, as $n \to \infty$,*

$$\mathrm{Bin}(n, p_n) \xrightarrow{d} \mathrm{Pois}(\lambda).$$

Proof. Using Remark 3.9 and Remark 3.10,

$$G_{\mathrm{Bin}(n,p_n)}(s) = \Big(1 + \frac{\lambda(s-1)}{n}\Big)^n \xrightarrow[n\to\infty]{} e^{\lambda(s-1)} = G_{\mathrm{Pois}(\lambda)}(s),$$

uniformly on $[0,1]$. Convergence of PGFs pointwise on $[0,1]$ implies convergence in distribution; see Feller [18, Ch. XI]. ∎

Remark 3.14 (Le Cam's inequality and rate of convergence)**.** A quantitative version of Theorem 3.13 is provided by *Le Cam's inequality* [30]: for independent Bernoulli variables $B_1, \dots, B_n$ with $\mathbb{P}(B_i = 1) = p_i$ and $S = \sum_i B_i$,

$$d_{\mathrm{TV}}(\mathcal{L}(S), \mathrm{Pois}(\textstyle\sum_i p_i)) \le \sum_{i=1}^n p_i^2.$$

In the homogeneous case $p_i = \lambda/n$ this gives $d_{\mathrm{TV}}(\mathrm{Bin}(n, \lambda/n), \mathrm{Pois}(\lambda)) \le \lambda^2/n$, so that the approximation is of order $1/n$. See also Barbour et al. [6] for a systematic treatment via Stein's method.

At the level of Galton–Watson processes, Theorem 3.13 implies that for every fixed generation $t \ge 0$ the law of Z_t under $\mathrm{GW}(\mathrm{Bin}(n, \lambda/n))$ converges to its law under $\mathrm{GW}(\mathrm{Pois}(\lambda))$. In fact this convergence extends to the total progeny T; see [41, Thm. 3.14]. This is the precise sense in which the Poisson Galton–Watson process is the correct limiting model for the exploration of $G(n, \lambda/n)$.

A further property of binomial offspring that is crucial for the stochastic domination bounds of Section 4.2 is the following elementary monotonicity. We first recall the relevant notion.

Definition 3.15 (Stochastic domination)**.** For two random variables X and Y taking values in $\mathbb{R}$, we say that X is *stochastically dominated* by Y, and we write $X \preceq_{\mathrm{st}} Y$, if

$$\mathbb{P}(X > t) \le \mathbb{P}(Y > t) \qquad \text{for every } t \in \mathbb{R}.$$

Proposition 3.16 (Monotonicity of binomials)**.** *Let $0 < m \le n$ be integers and $p \in [0,1]$. Then*

$$\mathrm{Bin}(m, p) \preceq_{\mathrm{st}} \mathrm{Bin}(n, p).$$

In particular, in the proof of the lower stochastic domination bound (Theorem 4.4), we compare the exploration process in $G(n,p)$ with a $\mathrm{GW}(\mathrm{Bin}(n-k, p))$ process. Since $n - k \le n$, Proposition 3.16 and the coupling above imply that the offspring at each step of the lower process are indeed dominated by the actual number of newly discovered vertices.

3.3.4 The conjugate parameter

The following parameter arises in the analysis of the supercritical regime and will be used in Section 4.4.7 (the duality principle).

Proposition 3.17 (Conjugate parameter)**.** *For every $\lambda > 1$ there exists a unique $\mu_\lambda \in (0,1)$ such that*

$$\mu_\lambda \, e^{-\mu_\lambda} \;=\; \lambda \, e^{-\lambda}. \tag{3.14}$$

It satisfies $\mu_\lambda = \lambda \, \eta_\lambda$, where η_λ is the extinction probability of the GW(Pois(λ)) *process.*

Proof. The function $\varphi(x) := xe^{-x}$ is continuous on $[0,\infty)$, strictly increasing on $[0,1]$ and strictly decreasing on $[1,\infty)$, with $\varphi(0) = 0$ and $\varphi(x) \to 0$ as $x \to \infty$. For every $\lambda > 1$, since $\varphi(\lambda) < \varphi(1) = e^{-1}$, the equation $\varphi(\mu) = \varphi(\lambda)$ has exactly one solution $\mu_\lambda \in (0,1)$.

To show $\mu_\lambda = \lambda\eta_\lambda$, recall from Proposition 3.11 that $\eta_\lambda = e^{\lambda(\eta_\lambda - 1)}$. Setting $\mu := \lambda\eta_\lambda$ and multiplying both sides by $\lambda e^{-\lambda}$ gives

$$\mu \, e^{-\mu} \;=\; \lambda \, \eta_\lambda \, e^{-\lambda\eta_\lambda} \;=\; \lambda \, e^{\lambda(\eta_\lambda - 1) - \lambda\eta_\lambda} \;=\; \lambda \, e^{-\lambda}.$$

Since $\eta_\lambda < 1$ we have $\mu < \lambda$, and one can verify from the fixed-point equation (3.12) that $\eta_\lambda < 1/\lambda$, so $\mu < 1$. Hence $\mu = \mu_\lambda$ by uniqueness. ∎

The conjugate parameter admits a direct probabilistic interpretation: *conditionally on extinction*, a supercritical GW(Pois(λ)) process has the same distribution as a subcritical GW(Pois(μ_λ)) process [41, Thm. 3.7]; this is a special case of a general duality principle for branching processes originally noted in Athreya and Ney [5]. At the random-graph level this corresponds to the *discrete duality principle* (Theorem 4.16): once the giant component is removed from $G(n, \lambda/n)$, the residual graph behaves like the subcritical random graph $G(\lceil n\eta_\lambda \rceil,\, \mu_\lambda/\lceil n\eta_\lambda \rceil)$. Since $\mu_\lambda < 1$, the results of Section 4.3 then give the size of the second-largest component.

Chapter 4

Phase Transition in Erdős–Rényi Graphs

Throughout this chapter we work in the sparse regime

$$p = \frac{\lambda}{n}, \qquad \lambda > 0 \text{ fixed.} \tag{4.1}$$

A central problem in the study of $G(n,p)$ is to understand the structure and size of its connected components as $n \to \infty$, in particular the emergence of a *giant component* [9, 19, 24]. Figure 4.1 illustrates the three main regimes of the model. In the subcritical regime ($\lambda < 1$) all connected components are typically small. Around the critical threshold ($\lambda \approx 1$) larger components of intermediate scale appear. In the supercritical regime ($\lambda > 1$) a giant component emerges, containing a positive fraction of the vertices, while the remaining components stay small.

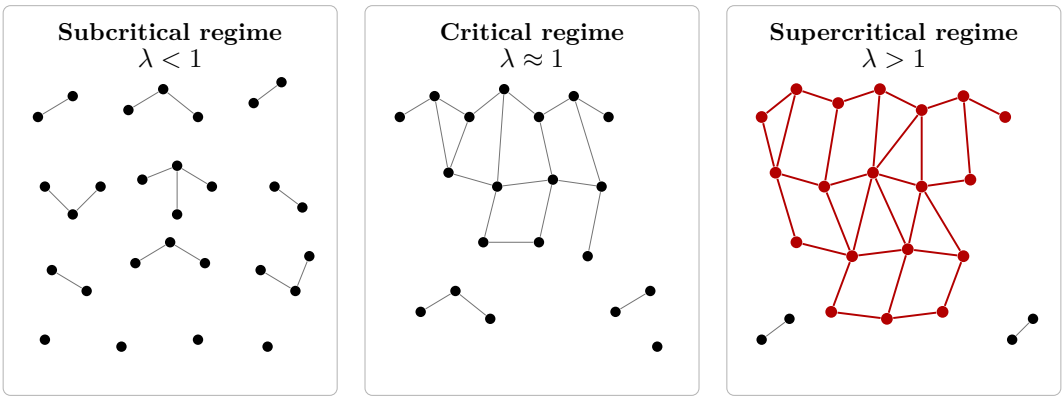


Figure 4.1: Schematic illustration of the phase transition in the sparse Erdős–Rényi random graph $G(n,p)$ with $p = \lambda/n$.

One of the key insights behind this analysis is that the local structure of $G(n,p)$ can be compared directly to suitable Galton–Watson branching processes [see 41, Sec. 4.1]. This connection becomes apparent when one explores the connected component of a given vertex v by a breadth–first search. Here, breadth–first search

means that the component is explored layer by layer: starting from v, one first reveals all neighbours of v, then the neighbours of those newly discovered vertices that have not yet been seen, and so on. Equivalently, vertices are explored in the same order in which they are discovered, so that the exploration naturally proceeds generation by generation [29, 36]. At each step of the exploration, the number of newly discovered vertices behaves, up to negligible correlations, like a binomial random variable. This observation allows one to compare the exploration process with branching processes having binomial or Poisson offspring distributions, introduced and studied in Section 3.3.

The aim of this chapter is to make this comparison precise and to use it as the main tool to analyse the phase transition of $G(n,p)$. We begin in Section 4.1 by explaining how the breadth–first exploration of a connected component can be compared heuristically with a Galton–Watson branching process. In Section 4.2 we make this comparison rigorous by proving upper and lower stochastic domination bounds for the size of a connected component in terms of the total progeny of suitable branching processes. We then apply these bounds to the different regimes of the Erdős–Rényi graph. In Section 4.3 we study the subcritical case $\lambda < 1$ and show that all components remain small, with the largest one of logarithmic order. In Section 4.4 we consider the supercritical case $\lambda > 1$, where the branching process survives with positive probability and a unique giant component of linear order emerges; for this regime we give a detailed, step–by–step proof of the law of large numbers for the size of the giant component (Theorem 4.8). Finally, in Section 4.5 we briefly discuss the critical window, where the branching–process approximation becomes more delicate and the largest components have size of order $n^{2/3}$. In this way, extinction and survival of branching processes provide the guiding principle for the phase transition in the Erdős–Rényi random graph, corresponding respectively to the absence or emergence of a giant component.

Remark 4.1 (Branching process viewpoint)**.** A convenient way to analyse the component structure of an Erdős–Rényi graph $G(n,p)$ is to explore the connected component of a vertex using a breadth–first search (BFS). During the initial stages of this exploration, most vertices are still unexplored. Thus, when exploring a vertex, each remaining vertex is connected to it independently with probability p, and the number of newly discovered neighbours is approximately distributed as $\mathrm{Bin}(n,p)$ with expectation close to np (see Remark 3.9). Consequently, the exploration behaves approximately like a Galton–Watson branching process [see, e.g., 8, 25].

Therefore, the exploration process can be approximated by a branching process whose mean offspring number satisfies

$$\mathbb{E}[\text{offspring}] \approx np.$$

This immediately suggests the existence of three distinct regimes, depending on the value of np:

Subcritical regime ($np < 1$) : the associated branching process becomes extinct

almost surely, and consequently all connected components of $G(n,p)$ remain small.

Critical regime ($np \approx 1$) : the branching process is critical, leading to unusually large but still sparse and tree like components.

Supercritical regime ($np > 1$) : the branching process has a positive probability of survival. In the random graph this corresponds to the sudden appearance of a unique *giant component*, whose size is linear in n.

In this way, the phase transition in the Erdős–Rényi graph mirrors the classical survival versus extinction transition of Galton–Watson branching processes (Theorem 3.3). In the remainder of this chapter, this correspondence will be made precise and used to derive rigorous results on the component structure of $G(n,p)$. Although we follow the branching–process approach as presented in van der Hofstad [41], the phase transition in Erdős–Rényi graphs was originally discovered by Erdős and Rényi using combinatorial methods [17], and subsequently refined by many authors including Bollobás [8], Łuczak et al. [33] and Pittel [37]. Since then, the emergence of the giant component and the critical behaviour of $G(n,p)$ have been extensively studied from several perspectives, including combinatorial, probabilistic, and analytical approaches; see for instance [7, 9, 15, 19, 24, 36, 39, 41]. A classical reference providing a detailed and self-contained treatment of the phase transition, including the giant component and the critical window, is Janson et al. [24]. The main results of this chapter are based on [41, Chapters 3–5], with particular emphasis on the comparison between random graphs and branching processes developed in Section 4.2 of that reference.

4.1 Comparison with Branching Processes

The main tool used to analyse the component structure of the Erdős–Rényi graph $G(n,p)$ is a precise comparison between the breadth–first exploration of a connected component and suitable Galton–Watson branching processes. We follow the approach described in [41, Section 4.2], which allows one to translate questions about cluster sizes in random graphs into questions about the total progeny of branching processes [see also 19, Ch. 5].

The starting point is the exploration of the connected component containing a fixed vertex $v \in [n]$. Vertices are revealed one by one in a breadth–first manner, keeping track of *active* vertices (whose neighbours are yet to be explored) and *neutral* vertices (not yet discovered). The exploration ends when no active vertices remain, at which point the entire connected component has been revealed.

Although the exploration process resembles a branching process, it is important to stress that it is *not* itself a Galton–Watson process. Indeed, the number of newly discovered vertices at each step depends on how many vertices have already been explored. As a consequence, the offspring numbers:

- are not independent,
- are not identically distributed,
- depend on the past of the process.

More precisely, when exploring the i-th vertex, the number of new neighbours is distributed as a binomial random variable whose number of trials equals the number of currently neutral vertices, which decreases over time. This prevents a direct identification of the exploration with a branching process. This dependence is often described as a depletion of neutral vertices effect, and motivates coupling and domination arguments; see Janson et al. [24, Ch. 4], Durrett [15, Ch. 6], and the general framework of Lindvall [31].

Despite these difficulties, by means of stochastic domination one can bound the exploration process from above and from below by two genuine Galton–Watson branching processes (cf. Proposition 3.16 and the general theory in Shaked and Shanthikumar 38); see van der Hofstad [41, Sec. 4.2] and Lyons and Peres [34]. The key observation is that, during the early stages of the exploration, only a small number of vertices has been discovered. Consequently, most vertices are still neutral, and their number remains close to n, so that the depletion effect is negligible. When exploring a given vertex, each neutral vertex is connected to it independently with probability p. Therefore, the number of newly discovered neighbours is distributed as $\mathrm{Bin}(N_t, p)$, where N_t is the number of neutral vertices at that stage. Since N_t is still close to n in the early phase, this can be approximated by $\mathrm{Bin}(n, p)$. Moreover, because only a small number of vertices has been explored, the set of neutral vertices does not change significantly from one step to the next, so the dependence between successive increments is negligible. Concretely, one constructs:

- an *upper* branching process whose offspring distribution dominates the number of newly discovered neighbours at each exploration step, and
- a *lower* branching process whose offspring distribution is dominated by the exploration increments, as long as the component size remains below a prescribed threshold.

The total progeny of these two branching processes, denoted respectively by $T^{\geq}$ and $T^{\leq}$, provide upper and lower bounds on the tail probabilities of the component size of the component containing v, namely, $|\mathcal{C}(v)|$.

In the sparse regime (4.1), the degree of a fixed vertex converges in distribution to a Poisson random variable (Theorem 3.13):

$$\mathrm{Bin}(n-1, p) \xrightarrow[n\to\infty]{d} \mathrm{Pois}(\lambda).$$

Hence, as long as the number of explored vertices is negligible compared to n, the exploration increments $\mathrm{Bin}(N_t, p)$ are well approximated by a $\mathrm{Pois}(\lambda)$ distribution. This explains why the Poisson approximation is valid during the initial phase of

the exploration. A quantitative control on this approximation is given by Le Cam's inequality (Remark 3.14). As the exploration progresses and a non-negligible fraction of vertices has been discovered, the number of remaining vertices decreases, and this approximation becomes less accurate.

This approximation naturally leads to a comparison with a Galton–Watson branching process with Poisson(λ) offspring distribution (Proposition 3.11). The qualitative behaviour of this branching process reflects the different regimes of the Erdős–Rényi graph. In particular, when $\lambda < 1$ the associated branching process becomes extinct almost surely, which corresponds to the fact that all connected components of $G(n,p)$ remain small. When $\lambda = 1$, the branching process is critical: it still becomes extinct almost surely, but can survive for a long time, leading to large but sparse components. When $\lambda > 1$, the branching process survives with positive probability ζ_λ, and this phenomenon is mirrored in the graph by the emergence of a unique giant component whose size is linear in n.

In the next section, this heuristic correspondence is made rigorous by means of stochastic domination bounds between the size of a connected component in $G(n,p)$ and the total progeny of suitable Galton–Watson branching processes.

4.2 Stochastic Domination of Cluster Sizes

In this section, we formalize the comparison between the breadth–first exploration of a connected component in the Erdős–Rényi graph $G(n,p)$ and Galton–Watson branching processes. This allows us to derive upper and lower bounds on component sizes via stochastic domination [38, Ch. 1].

Throughout this section we fix a vertex $v \in [n]$ and denote by $\mathcal{C}(v)$ the connected component containing v. We work in the sparse regime (4.1).

4.2.1 Exploration process

We explore the connected component $\mathcal{C}(v)$ of a fixed vertex $v \in [n]$ (for instance $v = 1$) by a breadth–first search. Let S_i denote the number of active vertices after i exploration steps, and let N_i denote the number of neutral vertices. Initially, the vertex v is declared active, while all other vertices are neutral. Thus,

$$S_0 = 1, \qquad N_0 = n - 1.$$

At each step $i \geq 1$, we select the first active vertex in the exploration queue (i.e. according to the breadth–first order), explore it, and then remove it from the set of active vertices, so that it becomes explored (inactive). During this step, we reveal all its neighbours that are still neutral, and declare them active. Let X_i denote the number of such newly discovered vertices.

The process then evolves according to the recursion

$$S_i = S_{i-1} + X_i - 1, \qquad N_i = n - i - S_i, \tag{4.2}$$

where the term -1 accounts for the fact that one active vertex is explored and removed from the queue at each step. The exploration terminates when no active vertices remain, that is, when all vertices in the connected component containing the initial vertex v have been discovered. This exploration recursion can be viewed as a random walk with increments $X_i - 1$; this encoding is standard in the study of random graphs, see e.g. Janson et al. [24, Ch. 4], Karp [25], and Frieze and Karoński [19, Ch. 5].
Define the stopping time

$$T := \min\{i \geq 1 : S_i = 0\}. \tag{4.3}$$

By construction,

$$|\mathcal{C}(v)| = T.$$

Before stating the stochastic domination results, we illustrate the exploration process with a simple example. The figure below shows the breadth–first exploration of the connected component of a fixed vertex together with the associated exploration process $(S_k)_{k\geq 0}$. In this way, the relation between the graph exploration, the random variables $(X_k)_{k\geq 1}$, and the process $(S_k)_{k\geq 0}$ becomes completely explicit. Figure 4.2 follows the breadth–first exploration of the connected component of vertex 1 all the way to its end. In panel (a), only the initial vertex is active, so $S_0 = 1$. In panel (b), vertex 1 is explored first. It has three neutral neighbours, namely vertices 2, 3, and 4. Thus

$$X_1 = 3, \qquad S_1 = S_0 + X_1 - 1 = 1 + 3 - 1 = 3.$$

So after the first exploration step there are three active vertices. In panel (c), the next active vertex is 2. It has exactly one neutral neighbour, namely vertex 5. Therefore

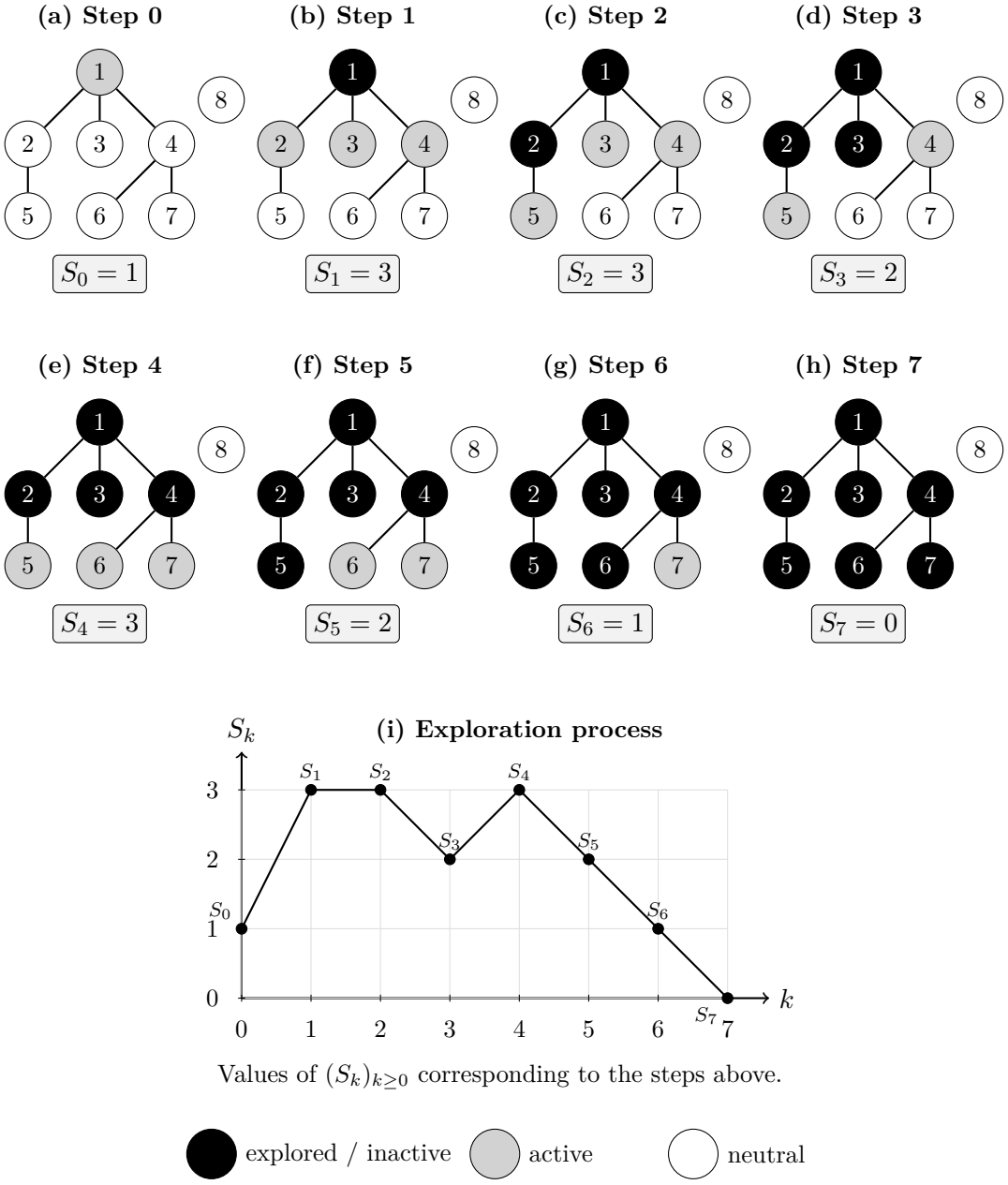
$$X_2 = 1, \qquad S_2 = S_1 + X_2 - 1 = 3 + 1 - 1 = 3.$$

Hence the number of active vertices stays unchanged. In panel (d), vertex 3 is explored and no new neutral neighbour is discovered, so that

$$X_3 = 0, \qquad S_3 = 2.$$

Continuing in the same way, in panel (e) vertex 4 produces two new vertices, so that $S_4 = 3$, while in panels (f) and (g) no new vertices are discovered, yielding $S_5 = 2$ and $S_6 = 1$. Finally, in panel (h), vertex 7 is explored and no new neighbour is found, so that

$$X_7 = 0, \qquad S_7 = S_6 + X_7 - 1 = 1 + 0 - 1 = 0.$$



Values of $(S_k)_{k\geq 0}$ corresponding to the steps above.

Figure 4.2: Breadth–first exploration of the connected component of vertex 1. Black vertices are explored, grey vertices are active, and white vertices are neutral. The lower panel shows the associated exploration process $(S_k)_{k\geq 0}$.

At this point no active vertices remain, so the exploration stops and the connected component of vertex 1 has been completely revealed. This also illustrates that the exploration process only reveals the connected component of the starting vertex, while other vertices of the graph (such as vertex 8) remain undiscovered.

The graph records how the number of active vertices evolves during the exploration. Each point corresponds to one exploration step. Starting from the initial vertex, the value increases when new neighbours are discovered, stays the same when exactly one new vertex replaces the explored one, and decreases when no new vertices are found.
In this example, the first step produces several new active vertices, so the value increases. In the next step, one new vertex is discovered, so the value remains unchanged. Afterwards, no new vertices are found, and the value starts to decrease. The exploration continues in this way until no active vertices remain, meaning that the whole connected component has been explored.

This random–walk representation is fundamental, as it makes explicit the connection between the exploration of components in $G(n,p)$ and Galton–Watson branching processes [25, 36]. When the variables (X_k) are approximated by independent Poisson random variables, the process behaves like the exploration of a branching process (see Section 3.3). This connection will be used in the following sections.

4.2.2 Upper bound: domination by a branching process

We first derive an upper bound on the size of $\mathcal{C}(v)$ by comparing the exploration process with a Galton–Watson branching process whose offspring distribution is $\mathrm{Bin}(n,p)$ (cf. Section 3.3.1). In other words, the exploration of a connected component can be stochastically dominated by the growth of a Galton–Watson tree with binomial offspring distribution. Indeed, when exploring a vertex in $G(n,p)$, some of its neighbours may have already been discovered, which reduces the number of new vertices that can be revealed. If we ignore this restriction and allow every potential neighbour to be available, the number of newly discovered vertices can only increase. This leads to a branching process that dominates the exploration process.

Theorem 4.2 (Stochastic domination of the cluster size [41, Thm. 4.2])**.** *For each $k \geq 1$,*

$$\mathbb{P}_{n,p}(|\mathcal{C}(v)| \geq k) \;\leq\; \mathbb{P}_{n,p}(T^{\geq} \geq k), \tag{4.4}$$

where $T^{\geq}$ is the total progeny of a binomial branching process with parameters (n,p), and $\mathbb{P}_{n,p}$ denotes the law of the Erdős–Rényi random graph $G(n,p)$.

Example 4.3. Fix $\lambda < 1$ and consider the Erdős–Rényi graph $G(n,p)$ with $p = \lambda/n$. Let v be a fixed vertex and let k be an integer such that $k \leq n$.

By Theorem 4.2, the event that the connected component of v has size at least k is dominated by the event that a subcritical Galton–Watson branching process produces at least k individuals:

$$\mathbb{P}\big(|\mathcal{C}(v)| \geq k\big) \;\leq\; \mathbb{P}\big(T^{\geq} \geq k\big).$$

Since the offspring mean is $\lambda < 1$, the branching process becomes extinct almost surely (Theorem 3.3) and its total progeny has an exponentially decaying tail

(Proposition 3.7). In particular, large values of k correspond to rare events in which the exploration process manages to sustain itself for an unusually long time.

From the graph perspective, this means that large connected components around a fixed vertex can only arise through atypical fluctuations of the exploration process, and their probability decreases exponentially with the component size.

Proof. Fix a vertex v (without loss of generality, $v = 1$) and explore its connected component by breadth–first search, using the notation introduced in Section 4.2. Thus S_i denotes the number of active vertices after i steps and N_i the number of neutral vertices, with $S_0 = 1$ and $N_0 = n - 1$.

The exploration evolves according to (4.2). In particular, at step $i \geq 1$ we process one active vertex and reveal its edges to the neutral vertices. Let

$$X_i := \#\{\text{neutral neighbours discovered at step } i\}.$$

Conditionally on N_{i-1}, each of the N_{i-1} neutral vertices is connected to the current vertex with probability p, independently of the others. Therefore

$$X_i \mid N_{i-1} \sim \mathrm{Bin}(N_{i-1}, p).$$

We define the stopping time T as in (4.3). By construction,

$$|\mathcal{C}(v)| = T.$$

We now construct an upper bounding process with independent increments. For each step i, define

$$Y_i := \#\{\text{neighbours of the current vertex that are not neutral at step } i\}.$$

Conditionally on N_{i-1}, each of the $n - N_{i-1}$ non-neutral vertices is connected to the current vertex independently with probability p. Hence

$$Y_i \mid N_{i-1} \sim \mathrm{Bin}(n - N_{i-1}, p),$$

and Y_i is independent of X_i.

Define

$$X_i^{\geq} := X_i + Y_i.$$

Then, conditionally on N_{i-1},

$$X_i^{\geq} \sim \mathrm{Bin}(n, p).$$

Moreover, since at each step all edges from the current vertex to the n vertices are revealed independently, it follows that the variables $(X_i^{\geq})_{i \geq 1}$ are independent and identically distributed. Finally,

$$X_i^{\geq} \geq X_i \qquad \text{almost surely.}$$

Define the corresponding upper process $(S_i^{\geq})$ by

$$S_0^{\geq} = 1, \qquad S_i^{\geq} = S_{i-1}^{\geq} + X_i^{\geq} - 1 \quad (i \geq 1).$$

Since $X_i^{\geq} \geq X_i$, it follows by induction that

$$S_i^{\geq} \geq S_i \qquad \text{for all } i \geq 0 \quad \text{almost surely.}$$

Since at each step we reveal all edges from the current vertex to the n vertices independently, each with probability p, it follows that the variables $(X_i^{\geq})_{i\geq 1}$ are independent and identically distributed with law $\mathrm{Bin}(n,p)$ (cf. Remark 3.9).

Therefore, the process $(S_i^{\geq})$ evolves as a random walk with i.i.d. increments $X_i^{\geq} - 1$, which is the standard representation of a Galton–Watson branching process with offspring distribution $\mathrm{Bin}(n,p)$ (see Chapter 3).

Define the stopping time

$$T^{\geq} := \min\{t \geq 1 : S_t^{\geq} = 0\}.$$

By the branching-process interpretation, $T^{\geq}$ coincides with the *total progeny* of this Galton–Watson process (see Section 3.2).

For any $k \geq 1$, we have

$$\{|\mathcal{C}(v)| \geq k\} = \{S_t > 0 \text{ for all } t \leq k-1\}.$$

Since $S_t^{\geq} \geq S_t$ for all t, it follows that

$$\{S_t > 0 \text{ for all } t \leq k-1\} \subseteq \{S_t^{\geq} > 0 \text{ for all } t \leq k-1\}.$$

Hence

$$\{|\mathcal{C}(v)| \geq k\} \subseteq \{T^{\geq} \geq k\}.$$

Taking probabilities, we obtain

$$\mathbb{P}_{n,p}(|\mathcal{C}(v)| \geq k) \;\leq\; \mathbb{P}_{n,p}(T^{\geq} \geq k),$$

which concludes the proof. ■

4.2.3 Lower bound: domination from below

We now derive a complementary lower bound. Before the exploration reaches size k, at most $k-1$ vertices have been discovered. Consequently, as long as $i \leq k-1$, the number of neutral vertices satisfies

$$N_{i-1} \geq n - k.$$

This observation, combined with the monotonicity of binomials (Proposition 3.16), allows us to compare the exploration process from below with a branching process whose offspring distribution is $\mathrm{Bin}(n-k,p)$.

Theorem 4.4 (Lower stochastic domination [41, Thm. 4.3])**.** *Let $k \in [n]$. Then*

$$\mathbb{P}_{n,p}(|\mathcal{C}(v)| \geq k) \;\geq\; \mathbb{P}_{n-k,p}(T^{\leq} \geq k),$$

where $T^{\leq}$ denotes the total progeny of a Galton–Watson branching process with offspring distribution $\mathrm{Bin}(n-k,p)$.

Proof. We use the same breadth–first exploration process as in Theorem 4.2. Let $v \in [n]$ be fixed. Since the Erdős–Rényi graph $G(n,p)$ is invariant under relabelling of the vertices, the distribution of $|\mathcal{C}(v)|$ does not depend on the choice of v. We may therefore, without loss of generality, take $v = 1$.

We denote by S_i, N_i and X_i the exploration process introduced above, which evolves according to (4.2). The size of the component containing 1 is given by the hitting time of 0. Indeed, by construction, the process S_t keeps track of the number of active vertices in the exploration: it starts from $S_0 = 1$, and at each step we explore one active vertex, remove it, and add its previously undiscovered neighbours. Thus $S_t > 0$ as long as there remain vertices to be explored, and the exploration stops exactly when no active vertices are left, that is, when $S_t = 0$. Since one vertex is explored at each step, the total number of steps until this time coincides with the size of the connected component. Hence

$$|\mathcal{C}(1)| = \min\{t \geq 1 : S_t = 0\}.$$

We define the stopping time

$$T_k := \min\{t \geq 0 : N_t \leq n-k\},$$

i.e. the first time that at most $n-k$ vertices are still neutral. If $|\mathcal{C}(1)| \geq k$, then by time k we have explored at least k vertices, and hence at time $k-1$ there are at most $n-k$ neutral vertices. Therefore $T_k \leq k-1$ on the event $\{|\mathcal{C}(1)| \geq k\}$. In terms of T_k we can write

$$\mathbb{P}_{n,p}(|\mathcal{C}(1)| \geq k) = \mathbb{P}_{n,p}(S_t > 0 \;\forall t \leq T_k). \tag{4.5}$$

For $i \geq 1$ let $(X_i^{\leq})_{i\geq 1}$ be an i.i.d. sequence with

$$X_i^{\leq} \sim \mathrm{Bin}(n-k,p),$$

independent of everything else. For $i \leq T_k$ define an auxiliary random variable

$$Y_i \sim \mathrm{Bin}\big(N_{i-1} - (n-k),\, p\big),$$

conditionally on N_{i-1}, and independent of $X_i^{\leq}$ given N_{i-1}. Set

$$X_i = X_i^{\leq} + Y_i, \qquad i \leq T_k.$$

By construction, conditionally on N_{i-1} we have

$$X_i^{\leq} \sim \mathrm{Bin}(n-k,p), \quad Y_i \sim \mathrm{Bin}\big(N_{i-1} - (n-k), p\big),$$

and these variables are independent given N_{i-1}. Hence

$$X_i = X_i^{\leq} + Y_i \sim \mathrm{Bin}(N_{i-1}, p),$$

as required for the exploration process. Moreover $Y_i \geq 0$ a.s., hence

$$X_i \;\geq\; X_i^{\leq} \qquad \text{a.s. for all } i \leq T_k. \tag{4.6}$$

Now define the *dominated* process

$$S_0^{\leq} = 1, \qquad S_i^{\leq} = S_{i-1}^{\leq} + X_i^{\leq} - 1, \quad i \geq 1.$$

This is exactly the exploration random walk associated with a Galton–Watson branching process with offspring distribution $\mathrm{Bin}(n-k, p)$. Let

$$T^{\leq} := \min\{t \geq 1 : S_t^{\leq} = 0\}$$

denote its total progeny (see Chapter 3).

From (4.6) and the recursion for S_i and $S_i^{\leq}$ it follows by induction that

$$S_i^{\leq} \;\leq\; S_i \qquad \text{for all } i \leq T_k \quad \text{a.s.} \tag{4.7}$$

If $S_t^{\leq} > 0$ for all $t \leq T_k$, then by (4.7) we also have $S_t > 0$ for all $t \leq T_k$. Hence

$$\{S_t^{\leq} > 0 \;\forall t \leq T_k\} \;\subseteq\; \{S_t > 0 \;\forall t \leq T_k\}.$$

Making use of (4.5) we obtain

$$\mathbb{P}_{n,p}(|\mathcal{C}(1)| \geq k) = \mathbb{P}_{n,p}(S_t > 0 \;\forall t \leq T_k) \;\geq\; \mathbb{P}_{n,p}(S_t^{\leq} > 0 \;\forall t \leq T_k).$$

Since $T_k \leq k-1$ whenever $|\mathcal{C}(1)| \geq k$, we may further bound

$$\mathbb{P}_{n,p}(S_t^{\leq} > 0 \;\forall t \leq T_k) \;\geq\; \mathbb{P}_{n,p}(S_t^{\leq} > 0 \;\forall t \leq k-1) = \mathbb{P}_{n-k,p}(T^{\leq} \geq k),$$

because $T^{\leq} = \min\{t : S_t^{\leq} = 0\}$. Combining the last two displays proves (4.4). ■

Theorems 4.2 and 4.4 show that the size of a connected component in $G(n,p)$ can be bounded above and below by the total progenies of two Galton–Watson branching processes. These results provide the rigorous foundation for the branching–process heuristics introduced earlier. In the next sections, these stochastic domination bounds will be used to derive precise asymptotic results for the largest component in the subcritical and supercritical regimes.

4.3 Subcritical Regime ($\lambda < 1$)

We now analyse the Erdős–Rényi random graph $G(n,p)$ in the subcritical regime, corresponding to (4.1) with $\lambda < 1$. In this regime the graph consists exclusively of small connected components, and no giant component emerges. The branching–process comparison developed in the previous section provides a precise probabilistic explanation of this behaviour [9, 19].

Let v be a fixed vertex and consider the size of its connected component $\mathcal{C}(v)$. By Theorem 4.2, for every $k \geq 1$,

$$\mathbb{P}_{n,p}(|\mathcal{C}(v)| \geq k) \;\leq\; \mathbb{P}(T^{\geq} \geq k),$$

where $T^{\geq}$ is the total progeny of a Galton–Watson process with offspring distribution $\mathrm{Bin}(n,p)$. Since $p = \lambda/n$ and $\lambda < 1$, by the Poisson limit theorem (Theorem 3.13) the binomial offspring distribution converges to a Poisson distribution:

$$\mathrm{Bin}(n,p) \xrightarrow[n\to\infty]{d} \mathrm{Pois}(\lambda).$$

Thus, for large n, the total progeny $T^{\geq}$ behaves like the total progeny T of a subcritical Poisson Galton–Watson process. In particular, when $\lambda < 1$, the total progeny has an exponential tail; see, e.g., van der Hofstad [41, Chapter 3]. More precisely, by Proposition 3.7, there exists a constant $I_\lambda > 0$ such that

$$\mathbb{P}(T \geq k) \leq e^{-I_\lambda k}, \qquad \text{for all } k \geq 1.$$

Exponential tail estimates for component sizes in the subcritical regime go back to the classical theory, see e.g. Janson et al. [24, Ch. 4], Bollobás [9, Ch. 5], and Łuczak et al. [33].

The constant I_λ is given explicitly by

$$I_\lambda := \lambda - 1 - \log \lambda,$$

which satisfies $I_\lambda > 0$ for all $\lambda < 1$. This constant arises naturally from large deviation theory for Poisson random variables [11]; see also Durrett [15]. As discussed in Chapter 2 (Section 2.2.1), it coincides with the value of the associated rate function at $x = 1$, namely

$$I(1) = \lambda - 1 - \log \lambda.$$

This quantity governs the exponential decay of the probability that a subcritical branching process survives for an unusually long time.

4.3.1 Expected component size

The branching–process approximation also yields the correct order of magnitude for the expected component size.

More precisely, the exploration of $\mathcal{C}(v)$ can be approximated by a Galton–Watson process with offspring distribution converging to $\mathrm{Pois}(\lambda)$ when $p = \lambda/n$ (Theorem 3.13). Let T denote the total progeny of a Poisson Galton–Watson process with mean $\lambda < 1$. By Proposition 3.6,

$$\mathbb{E}[T] = \frac{1}{1-\lambda}.$$

To relate this to the random graph, consider the Galton–Watson process with offspring distribution $\mathrm{Bin}(n,p)$ arising from the exploration approximation, and let T_n denote its total progeny. Since $\mathrm{Bin}(n,p) \Rightarrow \mathrm{Pois}(\lambda)$ when $p = \lambda/n$, we have $T_n \Rightarrow T$. Moreover, in the subcritical regime the expectations remain uniformly bounded, so that $\mathbb{E}[T_n] \to \mathbb{E}[T]$.

Using the stochastic domination bounds from Section 4.2, one obtains the same asymptotic behaviour for the expected component size, namely

$$\mathbb{E}_{n,p}\big[|\mathcal{C}(v)|\big] \xrightarrow[n\to\infty]{} \mathbb{E}[T] = \frac{1}{1-\lambda}.$$

Moreover, by Proposition 3.7, the total progeny has an exponential tail in the subcritical regime. In particular, large components occur with exponentially small probability, which confirms that typical components remain small when $\lambda < 1$.

4.3.2 Size of the largest component

We now turn to the size of the largest connected component

$$C_{\max} := \max_{v\in[n]} |\mathcal{C}(v)|.$$

The exponential decay governed by I_λ determines the precise logarithmic order of $C_{\max}$ [see also 17, 33].

Theorem 4.5 (Upper bound on the largest subcritical component [41, Thm. 4.4])**.** *Let $\lambda < 1$. For every $a > 1/I_\lambda$ there exists $\delta > 0$ such that*

$$\mathbb{P}_{n,p}(C_{\max} \geq a \log n) = O(n^{-\delta}).$$

Theorem 4.6 (Lower bound on the largest subcritical component [41, Thm. 4.5])**.** *Let $\lambda < 1$. For every $a < 1/I_\lambda$ there exists $\delta > 0$ such that*

$$\mathbb{P}_{n,p}(C_{\max} \leq a \log n) = O(n^{-\delta}).$$

We now prove Theorems 4.5 and 4.6. The argument combines the exponential tail estimate of Proposition 3.7, a first moment bound via Markov's inequality (Appendix B), and a second moment bound (Appendix B); see also Alon and Spencer [4, Ch. 4] for a general treatment of the moment method in probabilistic combinatorics. A fully detailed treatment, including all technical estimates, can be found in van der Hofstad [41, Sections 4.3.2–4.3.3].

Proof of Theorems 4.5 and 4.6. We start from the exploration process $(S_k)_{k\geq 0}$ introduced in Section 4.2, with recursion $S_k = S_{k-1} + X_k - 1$ and hitting time $T = \min\{k \geq 1 : S_k = 0\}$, so that $|\mathcal{C}(v)| = T$ by construction.

By Theorems 4.2 and 4.4, the component size $|\mathcal{C}(v)|$ is sandwiched between the total progenies of two Galton–Watson processes with offspring distributions $\mathrm{Bin}(n,p)$ and $\mathrm{Bin}(n-k,p)$, both converging to $\mathrm{Pois}(\lambda)$ as $n \to \infty$ (Theorem 3.13). By Proposition 3.7, there exist constants $0 < c \leq C < \infty$ independent of n and k such that

$$c\, e^{-I_\lambda k} \;\leq\; \mathbb{P}(|\mathcal{C}(v)| \geq k) \;\leq\; C\, e^{-I_\lambda k}, \tag{4.8}$$

where $I_\lambda = \lambda - 1 - \log\lambda > 0$.

Define

$$Z_{\geq k} \;:=\; \sum_{v\in[n]} \mathbf{1}_{\{|\mathcal{C}(v)|\geq k\}},$$

the number of vertices in components of size at least k, so that $\{C_{\max} \geq k\} = \{Z_{\geq k} \geq k\}$. Since $G(n,p)$ is exchangeable in its vertex labels, $|\mathcal{C}(v)| \overset{d}{=} |\mathcal{C}(v')|$ for all $v \neq v'$, and by symmetry

$$c\, n\, e^{-I_\lambda k} \;\leq\; \mathbb{E}[Z_{\geq k}] \;=\; n\,\mathbb{P}(|\mathcal{C}(1)| \geq k) \;\leq\; C\, n\, e^{-I_\lambda k}. \tag{4.9}$$

Upper bound (Theorem 4.5). Let $k = a\log n$ with $a > 1/I_\lambda$, so that $1 - aI_\lambda < 0$. From the upper bound in (4.9),

$$\mathbb{E}[Z_{\geq k}] \;\leq\; C\, n^{1-aI_\lambda} \;\to\; 0.$$

By Markov's inequality,

$$\mathbb{P}(C_{\max} \geq k) \;=\; \mathbb{P}(Z_{\geq k} \geq k) \;\leq\; \frac{C\, n^{1-aI_\lambda}}{a\log n} \;=\; O(n^{-\delta}),$$

with $\delta = aI_\lambda - 1 > 0$, which proves Theorem 4.5.

Lower bound (Theorem 4.6). Let $k = a\log n$ with $a < 1/I_\lambda$, so that $1 - aI_\lambda > 0$. From the lower bound in (4.9),

$$\mathbb{E}[Z_{\geq k}] \;\geq\; c\, n^{1-aI_\lambda} \;\to\; \infty.$$

To show that $Z_{\geq k} > 0$ with high probability, we use the second moment method (Proposition B.4; see also Alon and Spencer 4, Ch. 4):

$$\mathbb{P}(Z_{\geq k} > 0) \;\geq\; \frac{\big(\mathbb{E}[Z_{\geq k}]\big)^2}{\mathbb{E}[Z_{\geq k}^2]}.$$

We bound the second moment by expanding

$$\mathbb{E}[Z_{\geq k}^2] \;=\; \sum_{v,v'\in[n]} \mathbb{P}\big(|\mathcal{C}(v)| \geq k,\; |\mathcal{C}(v')| \geq k\big).$$

The n diagonal terms ($v = v'$) each contribute at most $Ce^{-I_\lambda k}$, giving a total of at most $Cne^{-I_\lambda k} \le C\,\mathbb{E}[Z_{\ge k}]$. For the off-diagonal terms ($v \neq v'$), note that if v and v' are in different components their events are independent, while if they are in the same component both indicators equal 1 together; in either case the joint probability is at most $Ce^{-I_\lambda k}$ per vertex, so

$$\mathbb{E}[Z_{\ge k}^2] \;\le\; C\,\mathbb{E}[Z_{\ge k}] \;+\; n^2C^2e^{-2I_\lambda k} \;\le\; C\,\mathbb{E}[Z_{\ge k}] \;+\; \frac{C^2}{c^2}\big(\mathbb{E}[Z_{\ge k}]\big)^2.$$

Therefore

$$\begin{aligned}\mathbb{P}(C_{\max} \le a\log n) \;&=\; \mathbb{P}(Z_{\ge k} = 0) \;\le\; 1 - \frac{\big(\mathbb{E}[Z_{\ge k}]\big)^2}{\mathbb{E}[Z_{\ge k}^2]} \\ &\le\; \frac{C}{\mathbb{E}[Z_{\ge k}]} \;\le\; \frac{C}{c\,n^{1-aI_\lambda}} \;=\; O(n^{-\delta}),\end{aligned}$$

with $\delta = 1 - aI_\lambda > 0$, which proves Theorem 4.6.

■

The upper bound relies on a first moment argument together with stochastic domination, while the lower bound requires a second moment argument to control dependencies between different vertices.

Combining Theorems 4.5 and 4.6 yields the following law of large logarithm.

Corollary 4.7. *If $\lambda < 1$, then*

$$\lim_{n\to\infty} \frac{C_{\max}}{\log n} = \frac{1}{I_\lambda} \quad \textit{in probability.}$$

Proof. Let $\varepsilon > 0$ be arbitrary, and set

$$a_- := \frac{1}{I_\lambda} - \varepsilon, \qquad a_+ := \frac{1}{I_\lambda} + \varepsilon.$$

For ε small enough we have

$$a_- < \frac{1}{I_\lambda} < a_+.$$

By Theorems 4.5 and 4.6, there exist $\delta_1, \delta_2 > 0$ such that

$$\mathbb{P}_{n,p}(C_{\max} \ge a_+\log n) = O(n^{-\delta_1}), \qquad \mathbb{P}_{n,p}(C_{\max} \le a_-\log n) = O(n^{-\delta_2}).$$

Therefore,

$$\begin{aligned}\mathbb{P}_{n,p}\left(\left|\frac{C_{\max}}{\log n} - \frac{1}{I_\lambda}\right| \ge \varepsilon\right) &\le \mathbb{P}_{n,p}(C_{\max} \le a_-\log n) + \mathbb{P}_{n,p}(C_{\max} \ge a_+\log n) \\ &= O(n^{-\delta_1}) + O(n^{-\delta_2}) = O\big(n^{-(\delta_1\wedge\delta_2)}\big),\end{aligned}$$

from which the result follows. ■

At a heuristic level, the upper bound follows by applying a union bound over all vertices together with the exponential tail of subcritical branching processes, while the lower bound is obtained by showing that with high probability at least one vertex initiates a component of size comparable to $(\log n)/I_\lambda$.

4.3.3 Interpretation

The subcritical Erdős–Rényi graph consists entirely of small, tree–like components. The branching–process comparison explains this phenomenon in a transparent way: the associated Galton–Watson process is subcritical and therefore becomes extinct almost surely (Theorem 3.3).

Large components correspond to rare large deviations of the branching process, and their probability decays exponentially with rate I_λ. The largest component is obtained by balancing this exponential decay against the number of vertices n, leading to a logarithmic scaling of order $(\log n)/I_\lambda$.

4.4 Supercritical Regime $(\lambda > 1)$

We now consider the Erdős–Rényi random graph $G(n,p)$ in the *supercritical regime*, where

$$p = \frac{\lambda}{n}, \qquad \lambda > 1.$$

In this regime the graph undergoes a fundamental structural change: with high probability a connected component containing a positive fraction of all vertices emerges. This is the appearance of the *giant component* [8, 17, 39]. Figure 4.1 gives a visual impression of how the giant component dominates the graph structure when $\lambda > 1$.

As in the subcritical case, the main tool is the comparison between the breadth–first exploration of a connected component in $G(n,p)$ and suitable Galton–Watson branching processes, as developed in Section 4.2. In the supercritical regime, this comparison links the emergence of the giant component to the *survival* of a supercritical branching process, in contrast to the subcritical case where extinction governed the absence of large components.

The key quantity governing the size of the giant component is the survival probability of the Poisson(λ) Galton–Watson branching process, which was studied in detail in Section 3.3.2 of Chapter 3. Recall from Proposition 3.11 that when $\lambda > 1$ the branching process survives with positive probability, and from equation (3.5) that the survival probability is

$$\zeta_\lambda \;:=\; 1 - \eta_\lambda \;>\; 0,$$

where η_λ is the extinction probability, characterized as the smallest solution of $\eta = e^{-\lambda(1-\eta)}$ (cf. equation (3.12)). The behaviour of ζ_λ as a function of λ is illustrated in Figure 3.2: it increases continuously from 0 as λ crosses 1 from above, and tends to 1 as $\lambda \to \infty$.

The relevance of ζ_λ for $G(n,p)$ comes from the branching-process approximation of the exploration process. During the early stages of a breadth-first exploration starting from a vertex v, the number of newly discovered neighbours at each step is approximately Pois(λ), as argued in Section 4.1 and Theorem 3.13. If the associated branching process survives, the exploration continues and the component of v grows to linear size; if it becomes extinct, the component remains small. Therefore, a

uniformly chosen vertex belongs to a large connected component with probability approximately ζ_λ [41, Ch. 4]. The following theorem makes this precise.

Theorem 4.8 (Law of large numbers for the giant component [41, Thm. 4.8])**.** *Fix $\lambda > 1$ and let $p = \lambda/n$. Let $\zeta_\lambda = 1 - \eta_\lambda$ be the survival probability of the* Pois(λ) *Galton–Watson process, as defined in Chapter 3 (equation* (3.5)*). Then, for every $\nu \in (1/2, 1)$ there exists $\delta = \delta(\nu, \lambda) > 0$ such that*

$$\mathbb{P}_\lambda\Big(\big||\mathcal{C}_{\max}| - \zeta_\lambda n\big| \geq n^\nu\Big) \;=\; O(n^{-\delta}) \qquad (n \to \infty). \tag{4.10}$$

Interpretation. Theorem 4.8 states that the largest connected component has size sharply concentrated around $\zeta_\lambda n$, up to fluctuations of order n^ν for any $\nu \in (1/2, 1)$. In particular, a *unique* giant component of linear size emerges, while all remaining components are of logarithmic order in n [41, see the discussion following Thm. 4.8]. The fraction of vertices in the giant component converges to ζ_λ, which by Proposition 3.11 (and Figure 3.2) is a continuous and strictly increasing function of λ for $\lambda > 1$.

We now give a detailed proof of Theorem 4.8, following the four–step strategy of van der Hofstad [41, Sec. 4.4] but carrying out each step explicitly. The only result that we use without proof is the variance estimate of Proposition 4.10 (Step 2), whose proof is purely technical and can be found in van der Hofstad [41, Prop. 4.10]. Alternative approaches to the law of large numbers for the giant component can be found in Bollobás [8], Janson et al. [24, Ch. 5], and the short proof of Krivelevich and Sudakov [29].

4.4.1 Strategy of the proof

The proof is based on the analysis of the random variable

$$Z_{\geq k} \;:=\; \sum_{v \in [n]} \mathbf{1}_{\{|\mathcal{C}(v)| \geq k\}}, \tag{4.11}$$

which counts the number of vertices lying in connected components of size at least k. Throughout we take

$$k_n = K \log n,$$

with $K > 0$ a sufficiently large constant to be fixed later, and we recall that $I_\lambda = \lambda - 1 - \log\lambda$ denotes the large–deviation rate constant introduced in (4.8). The argument proceeds in four steps.

Step 1 (Expectation). We show that $\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n) = \zeta_\lambda + O(k_n/n)$, and consequently $\mathbb{E}_\lambda[Z_{\geq k_n}] = \zeta_\lambda n + O(k_n)$. The mass counted by $Z_{\geq k_n}$ is therefore concentrated, in expectation, around $\zeta_\lambda n$.

Step 2 (Variance and concentration). Using a variance estimate for $Z_{\geq k_n}$, we deduce via Chebyshev's inequality that $Z_{\geq k_n}$ is concentrated around its mean, i.e. $|Z_{\geq k_n} - \zeta_\lambda n| < n^\nu$ with high probability.

Step 3 (No middle ground). We prove that, with high probability, no connected component has size in the intermediate range $[k_n, \alpha n]$ for a suitable $\alpha < \zeta_\lambda$. Every component is therefore either small (of order $\log n$) or large (of order n).

Step 4 (Completion). Combining Steps 2 and 3, we show that all the mass counted by $Z_{\geq k_n}$ belongs to a single component, so that $|\mathcal{C}_{\max}| = Z_{\geq k_n}$ with high probability. The concentration of $Z_{\geq k_n}$ then yields (4.10).

4.4.2 Step 1: the cluster tail and the expected mass in large clusters

The first step relates the probability that a fixed vertex lies in a large component to the survival probability ζ_λ of the limiting Poisson branching process.

Proposition 4.9 (Cluster tail equals the survival probability [41, Prop. 4.9])**.** *Fix $\lambda > 1$ and let $k_n = \lceil a \log n \rceil$ with $a > 1/I_\lambda$. Then, for n sufficiently large,*

$$\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n) \;=\; \zeta_\lambda + O(k_n/n).$$

Proof. We prove matching upper and lower bounds.

Upper bound. By the stochastic domination of Theorem 4.2, $|\mathcal{C}(v)|$ is dominated by the total progeny $T^{\geq}$ of a binomial branching process with parameters $(n, \lambda/n)$. Comparing this binomial branching process with the Poisson(λ) branching process, whose total progeny we denote by T^*, and using the coupling between binomial and Poisson offspring (Theorem 3.13; see also van der Hofstad 41, Thm. 3.14), one obtains

$$\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n) \;\leq\; \mathbb{P}(T^* \geq k_n) + O(k_n/n). \tag{4.12}$$

The total progeny of the supercritical Poisson process is infinite exactly on the survival event, which has probability ζ_λ (Proposition 3.11). Splitting according to whether T^* is finite or infinite,

$$\mathbb{P}(T^* \geq k_n) = \mathbb{P}(T^* = \infty) + \mathbb{P}(k_n \leq T^* < \infty) = \zeta_\lambda + \mathbb{P}(k_n \leq T^* < \infty).$$

On the event $\{T^* < \infty\}$ the process is conditioned to die out, and its total progeny has an exponentially decaying tail: by Proposition 3.7 (see also van der Hofstad 41, Thm. 3.8),

$$\mathbb{P}(k_n \leq T^* < \infty) \;\leq\; e^{-I_\lambda k_n}.$$

Since $k_n \geq a \log n$ with $a > 1/I_\lambda$, we have $e^{-I_\lambda k_n} \leq n^{-aI_\lambda} = o(1/n)$, and therefore

$$\mathbb{P}(T^* \geq k_n) = \zeta_\lambda + O(e^{-I_\lambda k_n}) = \zeta_\lambda + o(1/n). \tag{4.13}$$

Substituting (4.13) into (4.12) gives the upper bound $\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n) \leq \zeta_\lambda + O(k_n/n)$.

Lower bound. By the lower domination of Theorem 4.4, $|\mathcal{C}(v)|$ dominates the total progeny of a binomial branching process with parameters $(n - k_n, \lambda/n)$, whose offspring mean is

$$\lambda_n := \lambda\Big(1 - \frac{k_n}{n}\Big).$$

Arguing as above, but now comparing with a Poisson(λ_n) branching process with total progeny $T^*_{\lambda_n}$,

$$\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n) \;\geq\; \mathbb{P}(T^*_{\lambda_n} \geq k_n) + O(k_n/n) = \zeta_{\lambda_n} + O(k_n/n), \tag{4.14}$$

where we used, exactly as in (4.13), that the finite part of the progeny contributes only $o(1/n)$. It remains to compare ζ_{λ_n} with ζ_λ. Since $\lambda > 1$, the survival probability $\zeta_\mu = 1 - \eta_\mu$ is differentiable in μ in a neighbourhood of λ (Proposition 3.11; see also van der Hofstad 41, Cor. 3.19). By the mean value theorem there exists $\lambda_n^* \in (\lambda_n, \lambda)$ such that

$$\eta_{\lambda_n} = \eta_\lambda + (\lambda_n - \lambda)\,\frac{d}{d\mu}\eta_\mu\Big|_{\mu=\lambda_n^*} = \eta_\lambda + O(k_n/n),$$

because $\lambda - \lambda_n = \lambda k_n/n = O(k_n/n)$ and the derivative is bounded near λ. Hence $\zeta_{\lambda_n} = \zeta_\lambda + O(k_n/n)$, and (4.14) yields the matching lower bound.

Combining the two bounds completes the proof. ∎

A direct consequence of Proposition 4.9 concerns the expected value of $Z_{\geq k_n}$. By the exchangeability of the vertices of $G(n,p)$,

$$\mathbb{E}_\lambda[Z_{\geq k_n}] \;=\; n\,\mathbb{P}_\lambda(|\mathcal{C}(1)| \geq k_n) \;=\; \zeta_\lambda n + O(k_n). \tag{4.15}$$

Thus, in expectation, the number of vertices in large components is $\zeta_\lambda n$ up to a correction of order $k_n = K \log n$.

4.4.3 Step 2: concentration of the mass in large clusters

We now show that $Z_{\geq k_n}$ is not only correct in expectation but in fact sharply concentrated around $\zeta_\lambda n$. This requires controlling its variance. In the statement we use the notation

$$\chi_{<k}(\lambda) \;:=\; \mathbb{E}_\lambda\big[|\mathcal{C}(v)|\,\mathbf{1}_{\{|\mathcal{C}(v)|<k\}}\big], \tag{4.16}$$

that is, the expected size of the component of v restricted to the event that this component is smaller than k.

Proposition 4.10 (Variance estimate for $Z_{\geq k}$ [41, Prop. 4.10])**.** *For every n and every $k \in [n]$,*

$$\mathrm{Var}_\lambda(Z_{\geq k}) \;\leq\; (\lambda k + 1)\, n\, \chi_{<k}(\lambda). \tag{4.17}$$

We do not reproduce the proof of Proposition 4.10, which is of a purely technical nature: it relies on a careful bookkeeping of the covariances $\mathrm{Cov}_\lambda(\mathbf{1}_{\{|\mathcal{C}(u)|\geq k\}}, \mathbf{1}_{\{|\mathcal{C}(v)|\geq k\}})$ obtained by re–running the exploration process from two distinct vertices and distinguishing whether or not they lie in the same component. The crucial feature of the bound (4.17) is the factor $\chi_{<k}(\lambda)$: in the supercritical regime the bulk of the mass sits in the giant component, which has size of order n and is therefore *excluded* from the truncated first moment (4.16). This makes $\chi_{<k}(\lambda)$ of order k rather than of order n, so that (4.17) is of order $k^2 n$ instead of the trivial order n^2. The complete argument is given in van der Hofstad [41, Prop. 4.10].

The variance estimate immediately yields the desired concentration.

Corollary 4.11 (Concentration of $Z_{\geq k_n}$ [41, Cor. 4.11])**.** *Fix $\nu \in (1/2, 1)$ and $k_n = K \log n$ with K sufficiently large. Then, for every $\delta < 2\nu - 1$, as $n \to \infty$,*

$$\mathbb{P}_\lambda\big(|Z_{\geq k_n} - \zeta_\lambda n| \geq n^\nu\big) \;=\; O(n^{-\delta}).$$

Proof. By (4.15) we have $\mathbb{E}_\lambda[Z_{\geq k_n}] = \zeta_\lambda n + O(k_n)$, and since $k_n = K \log n = o(n^\nu)$, for n large enough the deviation of $Z_{\geq k_n}$ from $\zeta_\lambda n$ by at least n^ν forces a deviation from its mean by at least $n^\nu/2$:

$$\big\{|Z_{\geq k_n} - \zeta_\lambda n| \geq n^\nu\big\} \;\subseteq\; \big\{|Z_{\geq k_n} - \mathbb{E}_\lambda[Z_{\geq k_n}]| \geq n^\nu/2\big\}.$$

To bound $\chi_{<k_n}(\lambda)$ we note that on the event $\{|\mathcal{C}(v)| < k_n\}$ the component size is at most k_n, so that $\chi_{<k_n}(\lambda) \leq k_n$. Hence, by Proposition 4.10,

$$\mathrm{Var}_\lambda(Z_{\geq k_n}) \;\leq\; (\lambda k_n + 1)\, n\, \chi_{<k_n}(\lambda) \;\leq\; (\lambda k_n + 1)\, n\, k_n \;=\; O(n\, k_n^2).$$

By Chebyshev's inequality,

$$\mathbb{P}_\lambda\big(|Z_{\geq k_n} - \zeta_\lambda n| \geq n^\nu\big) \;\leq\; \frac{4\,\mathrm{Var}_\lambda(Z_{\geq k_n})}{n^{2\nu}} \;=\; O\!\left(\frac{k_n^2}{n^{2\nu-1}}\right) \;=\; O(n^{-\delta})$$

for every $\delta < 2\nu - 1$, since $k_n = K \log n$ grows only logarithmically and $2\nu - 1 > 0$. This proves the claim. ∎

4.4.4 Step 3: no middle ground

The third step shows that, with high probability, $G(n,p)$ has no connected component of *intermediate* size, i.e. of size between $k_n = K \log n$ and αn for a suitable $\alpha < \zeta_\lambda$. Components are therefore either small (of logarithmic order) or large (of linear order), with nothing in between. This dichotomy is the graph counterpart of the survival/extinction dichotomy for branching processes.

The argument is based on an exact distributional identity for the exploration random walk $(S_t)_{t \geq 0}$ of Section 4.2. We use the *formal continuation* of the recursion (4.2) beyond the stopping time T, so that (S_t) is defined for all $t \in [n]$ (allowing S_t to become negative); this does not affect the law of $|\mathcal{C}(v)|$, since we only use the identity on the event $\{|\mathcal{C}(v)| \geq t\}$, where $S_t \geq 0$.

Lemma 4.12 (Law of the exploration process [41, Prop. 4.6])**.** *For every $t \in [n]$, with (S_t) defined by the formal continuation of (4.2) and $p = \lambda/n$,*

$$S_t + (t-1) \;\sim\; \mathrm{Bin}\big(n-1,\, 1-(1-p)^t\big). \tag{4.18}$$

Proof. Let $N_t = n - t - S_t$ denote the number of neutral vertices at time t, as in (4.2). It is equivalent, and more convenient, to show

$$N_t \;\sim\; \mathrm{Bin}\big(n-1, (1-p)^t\big), \tag{4.19}$$

since $X \sim \mathrm{Bin}(m,q)$ if and only if $m - X \sim \mathrm{Bin}(m, 1-q)$, and $S_t + (t-1) = (n-1) - N_t$.

We prove (4.19) by induction on t. Initially $N_0 = n-1$, which is the $(\mathrm{Bin}(n-1,1))$ case. Conditionally on N_{t-1}, the increment at step t is the number of neutral neighbours of the explored vertex, hence $X_t \sim \mathrm{Bin}(N_{t-1}, p)$, and therefore

$$N_t \;=\; N_{t-1} - X_t \;\sim\; \mathrm{Bin}(N_{t-1},\, 1-p) \quad \text{conditionally on } N_{t-1}.$$

Thus each of the $n-1$ vertices other than v remains neutral at step t if and only if it remained neutral at step $t-1$ and was not discovered at step t, an event of conditional probability $1-p$. By induction, each such vertex is neutral at time t independently with probability $(1-p)^t$, which is exactly (4.19). This proves the lemma. ∎

To quantify the exponential decay we introduce, for $\alpha \in (0,1)$, the functions

$$g(\alpha;\lambda) := \frac{1-e^{-\lambda\alpha}}{\alpha}, \qquad J(\alpha;\lambda) := I_{g(\alpha;\lambda)}, \qquad I_x := x - 1 - \log x, \tag{4.20}$$

where I_x is the Poisson large–deviation rate function already used in (4.8). The function $\alpha \mapsto g(\alpha;\lambda)$ is strictly decreasing, with $g(\alpha;\lambda) \to \lambda$ as $\alpha \downarrow 0$, and the survival probability ζ_λ is the unique solution in $(0,1]$ of $g(\alpha;\lambda) = 1$ (equivalently $1-e^{-\lambda\alpha} = \alpha$, i.e. $e^{-\lambda\zeta_\lambda} = \eta_\lambda$). Consequently $g(\alpha;\lambda) > 1$ for every $\alpha < \zeta_\lambda$, and since $I_x > 0$ for $x \neq 1$ and I_x is increasing for $x \geq 1$, we have $J(\alpha;\lambda) > 0$ for all $\alpha < \zeta_\lambda$.

Proposition 4.13 (Exponential bound for clusters smaller than $\zeta_\lambda n$ [41, Prop. 4.12])**.** *Fix $\lambda > 1$ and let $k_n \to \infty$ as $n \to \infty$. Then, for every $\alpha < \zeta_\lambda$,*

$$\mathbb{P}_\lambda\big(k_n \leq |\mathcal{C}(v)| \leq \alpha n\big) \;\leq\; \frac{e^{-k_n J(\alpha;\lambda)}}{1-e^{-J(\alpha;\lambda)}}. \tag{4.21}$$

Proof. Since $\{|\mathcal{C}(v)| = t\} \subseteq \{S_t = 0\}$, summing over the admissible values of t gives

$$\mathbb{P}_\lambda\big(k_n \leq |\mathcal{C}(v)| \leq \alpha n\big) = \sum_{t=k_n}^{\lfloor \alpha n \rfloor} \mathbb{P}_\lambda\big(|\mathcal{C}(v)| = t\big) \;\leq\; \sum_{t=k_n}^{\lfloor \alpha n \rfloor} \mathbb{P}_\lambda(S_t = 0).$$

By Lemma 4.12, $S_t = 0$ means $\mathrm{Bin}(n-1, 1-(1-p)^t) = t-1$. Using $1-p \leq e^{-p}$, the success probability satisfies $1-(1-p)^t \geq 1-e^{-pt} =: r_t$, and a binomial with larger success probability is stochastically larger; hence

$$\mathbb{P}_\lambda(S_t = 0) \;\leq\; \mathbb{P}\big(\mathrm{Bin}(n, r_t) \leq t\big).$$

We bound the last probability by a Chernoff estimate. For every $s > 0$, Markov's inequality applied to $e^{-s\,\mathrm{Bin}(n,r_t)}$ gives

$$\mathbb{P}\big(\mathrm{Bin}(n,r_t) \leq t\big) \;\leq\; e^{st}\,\big(1 - r_t(1-e^{-s})\big)^n \;\leq\; \exp\!\big(st - n\,r_t(1-e^{-s})\big),$$

using $1 - x \leq e^{-x}$. Writing $\beta = t/n$, we have $n\,r_t = n(1-e^{-\lambda\beta}) = t\,g(\beta;\lambda)$ by (4.20). The exponent $s \mapsto st - t\,g(\beta;\lambda)(1-e^{-s})$ is minimized at $s^* = \log g(\beta;\lambda)$, and substituting s^* yields

$$\mathbb{P}_\lambda(S_t = 0) \;\leq\; \exp\!\big(-t\,I_{g(\beta;\lambda)}\big) \;=\; \exp\!\big(-t\,J(\beta;\lambda)\big), \qquad \beta = t/n.$$

For $t \le \alpha n$ we have $\beta \le \alpha < \zeta_\lambda$; since $g(\cdot;\lambda)$ is decreasing and $x \mapsto I_x$ is increasing on $[1,\infty)$ with $g(\beta;\lambda) \ge g(\alpha;\lambda) > 1$, it follows that $J(\beta;\lambda) \ge J(\alpha;\lambda)$. Therefore $\mathbb{P}_\lambda(S_t = 0) \le e^{-tJ(\alpha;\lambda)}$, and summing the resulting geometric series from $t = k_n$ gives

$$\mathbb{P}_\lambda\big(k_n \le |\mathcal{C}(v)| \le \alpha n\big) \;\le\; \sum_{t \ge k_n} e^{-tJ(\alpha;\lambda)} \;=\; \frac{e^{-k_n J(\alpha;\lambda)}}{1 - e^{-J(\alpha;\lambda)}},$$

which is (4.21). ∎

Corollary 4.14 (No intermediate clusters [41, Cor. 4.13])**.** *Fix $k_n = K \log n$ and $\alpha < \zeta_\lambda$, and set $\delta = K\,J(\alpha;\lambda) - 1$, which is positive for K sufficiently large. Then, with probability at least $1 - O(n^{-\delta})$, there is no connected component with size in the range $[k_n, \alpha n]$.*

Proof. By the union bound and exchangeability of the vertices,

$$\mathbb{P}_\lambda\big(\exists\, v : k_n \le |\mathcal{C}(v)| \le \alpha n\big) \;\le\; n\,\mathbb{P}_\lambda\big(k_n \le |\mathcal{C}(1)| \le \alpha n\big) \;\le\; \frac{n\, e^{-k_n J(\alpha;\lambda)}}{1 - e^{-J(\alpha;\lambda)}},$$

by Proposition 4.13. With $k_n = K \log n$ the right–hand side equals $C\, n^{1 - K J(\alpha;\lambda)} = O(n^{-\delta})$, where $C = 1/(1 - e^{-J(\alpha;\lambda)}) < \infty$ and $\delta = K J(\alpha;\lambda) - 1 > 0$ for K large enough. This proves the corollary. ∎

4.4.5 Step 4: completion of the proof

We now combine the previous steps. Fix $\nu \in (1/2, 1)$, $\alpha \in (\zeta_\lambda/2, \zeta_\lambda)$, and $k_n = K \log n$ with K sufficiently large. Let E_n be the event that

(1) $|Z_{\ge k_n} - \zeta_\lambda n| < n^\nu$; and

(2) there is no vertex $v \in [n]$ with $k_n \le |\mathcal{C}(v)| \le \alpha n$.

Lemma 4.15 ($|\mathcal{C}_{\max}|$ equals $Z_{\ge k_n}$ [41, Lem. 4.14])**.** *Fix $\alpha \in (\zeta_\lambda/2, \zeta_\lambda)$. Then E_n occurs with high probability: taking δ to be the minimum of the exponents in Corollaries 4.11 and 4.14, one has $\mathbb{P}_\lambda(E_n^c) = O(n^{-\delta})$. Moreover, on the event E_n,*

$$|\mathcal{C}_{\max}| = Z_{\ge k_n}.$$

Proof. The complement E_n^c is the union of the complements of (1) and (2), so by the union bound

$$\mathbb{P}_\lambda(E_n^c) \;\le\; \mathbb{P}_\lambda\big(|Z_{\ge k_n} - \zeta_\lambda n| \ge n^\nu\big) + \mathbb{P}_\lambda\big(\exists\, v : k_n \le |\mathcal{C}(v)| \le \alpha n\big).$$

By Corollary 4.11 the first term is $O(n^{-\delta})$, and by Corollary 4.14 the second term is $O(n^{-\delta})$; hence $\mathbb{P}_\lambda(E_n^c) = O(n^{-\delta})$ for K large enough.

It remains to prove that $|\mathcal{C}_{\max}| = Z_{\ge k_n}$ on E_n. On E_n, condition (1) gives $Z_{\ge k_n} \ge \zeta_\lambda n - n^\nu \ge 1$, so at least one component has size at least k_n; since $\mathcal{C}_{\max}$ is one of the components counted by $Z_{\ge k_n}$, we always have $|\mathcal{C}_{\max}| \le Z_{\ge k_n}$. Suppose, for contradiction, that $|\mathcal{C}_{\max}| < Z_{\ge k_n}$. Then the vertices counted by $Z_{\ge k_n}$ cannot

all belong to $\mathcal{C}_{\max}$, so there exist at least two distinct components of size at least k_n. By condition (2) no component has size in $[k_n, \alpha n]$, so both of these components have size at least αn. Consequently

$$Z_{\geq k_n} \;\geq\; 2\alpha n.$$

Since $\alpha > \zeta_\lambda/2$, we have $2\alpha > \zeta_\lambda$, and therefore $2\alpha n > \zeta_\lambda n + n^\nu$ for n large enough, contradicting condition (1), which gives $Z_{\geq k_n} < \zeta_\lambda n + n^\nu$. Hence $|\mathcal{C}_{\max}| = Z_{\geq k_n}$ on E_n. ∎

We can now complete the proof of the law of large numbers.

Proof of Theorem 4.8. Let E_n be the event defined above, with $\alpha \in (\zeta_\lambda/2, \zeta_\lambda)$ and $k_n = K \log n$ for K large. By Lemma 4.15, on E_n we have $|\mathcal{C}_{\max}| = Z_{\geq k_n}$, and by condition (1) of E_n also $|Z_{\geq k_n} - \zeta_\lambda n| < n^\nu$. Therefore, on E_n,

$$\big||\mathcal{C}_{\max}| - \zeta_\lambda n\big| = \big|Z_{\geq k_n} - \zeta_\lambda n\big| < n^\nu,$$

so that the event $\{\, ||\mathcal{C}_{\max}| - \zeta_\lambda n| \geq n^\nu \,\}$ is contained in E_n^c. By Lemma 4.15,

$$\mathbb{P}_\lambda\Big(\big||\mathcal{C}_{\max}| - \zeta_\lambda n\big| \geq n^\nu\Big) \;\leq\; \mathbb{P}_\lambda(E_n^c) \;=\; O(n^{-\delta}),$$

which is precisely (4.10). This completes the proof. ∎

4.4.6 Uniqueness of the giant component

The proof of Theorem 4.8 yields, as a by–product, the *uniqueness* of the giant component. Indeed, Lemma 4.15 shows that, with high probability, *all* the vertices counted by $Z_{\geq k_n}$, that is, all vertices in components of size at least $k_n = K \log n$, belong to a single connected component, namely $\mathcal{C}_{\max}$. Were the mass $\zeta_\lambda n$ split between two or more components of linear size, the total count $Z_{\geq k_n}$ would exceed $\zeta_\lambda n + n^\nu$, contradicting the concentration of Corollary 4.11; the absence of intermediate components (Corollary 4.14, Step 3) rules out the remaining possibility of a component of sublinear but non–logarithmic size.

Hence, in the supercritical regime, $G(n, \lambda/n)$ has, with high probability, a unique giant component of size $\zeta_\lambda n(1+o(1))$, while every other component has size $O(\log n)$. An alternative and classical route to uniqueness is the *sprinkling* argument of Ajtai et al. [1], in which a small fraction of the edges is reserved to connect any two large components with high probability; see also Bollobás [9, Ch. 5] and Janson et al. [24, Sec. 5.2].

4.4.7 The duality principle

Once the giant component has been identified, it is natural to ask about the structure of the remaining graph. A remarkable fact is that, after removing $\mathcal{C}_{\max}$, the residual graph behaves like a *subcritical* Erdős–Rényi graph.

For $\lambda > 1$, recall from Proposition 3.17 the definition of the *conjugate parameter* $\mu_\lambda \in (0,1)$ as the unique solution of

$$\mu_\lambda e^{-\mu_\lambda} = \lambda e^{-\lambda}, \tag{4.22}$$

with the equivalent representation $\mu_\lambda = \lambda\eta_\lambda$, where η_λ denotes the extinction probability of the Poisson(λ) Galton–Watson process.

The relevance of μ_λ comes from the branching-process approximation: components that do not grow to linear size correspond to realizations of the branching process that become extinct, and their behaviour is governed by a subcritical Poisson(μ_λ) process, as explained after Proposition 3.17.

Theorem 4.16 (Discrete duality principle [41, Thm. 4.15])**.** *Let $\lambda > 1$ and let $\mu_\lambda \in (0,1)$ be defined by (4.22). Set $m = \lceil n\eta_\lambda \rceil$. Then the vector of component sizes of $G(n, \lambda/n)$ with the giant component removed is close in law to that of the subcritical random graph $G(m, \mu_\lambda/m)$.*

Proof (sketch). We follow the heuristic argument of van der Hofstad [41, Sec. 4.4.6]; a fully rigorous proof is given in Janson et al. [24, Sec. 5.6] and Bollobás [9, Sec. 6.3]. By Theorem 4.8, the giant component contains $\zeta_\lambda n(1+o(1))$ vertices, so the number of vertices outside it is $m = n - |\mathcal{C}_{\max}| = \eta_\lambda n(1+o(1))$, in agreement with $m = \lceil n\eta_\lambda \rceil$.

Condition on the vertex set of the giant component and on its size $|\mathcal{C}_{\max}| = n-m$. The edges among the remaining m vertices were *not* examined when revealing $\mathcal{C}_{\max}$; conditionally on the removal, they are still present independently, each with the original probability $p = \lambda/n$. Hence the residual graph is an Erdős–Rényi graph on m vertices with edge probability λ/n. Writing this in terms of m, and using $m \approx \eta_\lambda n$ together with the identity $\lambda\eta_\lambda = \mu_\lambda$ from (4.22), the effective edge probability is

$$\frac{\lambda}{n} \;=\; \frac{\lambda m}{n} \cdot \frac{1}{m} \;\approx\; \frac{\lambda\eta_\lambda}{1} \cdot \frac{1}{m} \;=\; \frac{\mu_\lambda}{m}.$$

Thus the residual graph is, to leading order, distributed as $G(m, \mu_\lambda/m)$. Since $\mu_\lambda < 1$, this graph is subcritical, which is the content of the duality principle. We omit the estimates needed to control the conditioning and to make the phrase "close in law" precise, referring to the references above for the details. ∎

Since $\mu_\lambda < 1$, the residual graph is subcritical. As a consequence, the size of the second largest component satisfies

$$\frac{|\mathcal{C}^{(2)}|}{\log n} \;\xrightarrow{\;\mathbb{P}\;}\; \frac{1}{I_{\mu_\lambda}},$$

by the results of Section 4.3 applied with λ replaced by μ_λ.

4.4.8 Fluctuations of the giant component

The law of large numbers (Theorem 4.8) shows that $|\mathcal{C}_{\max}| = \zeta_\lambda n + o(n)$. A natural refinement is to study the fluctuations around this deterministic limit [10, 37, 40].

Theorem 4.17 (Central limit theorem for the giant component [41, Thm. 4.16])**.** *Fix $\lambda > 1$. Then*

$$\frac{|\mathcal{C}_{\max}| - \zeta_\lambda n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma_\lambda^2),$$

where

$$\sigma_\lambda^2 = \frac{\zeta_\lambda(1-\zeta_\lambda)}{(1-\lambda+\lambda\zeta_\lambda)^2}. \tag{4.23}$$

Interpretation. The fluctuations are of order $\sqrt{n}$ and asymptotically Gaussian. The factor $(1-\lambda+\lambda\zeta_\lambda)^{-2}$ reflects the dependence between vertices induced by the exploration process, in contrast with the variance $n\zeta_\lambda(1-\zeta_\lambda)$ that would arise under independence.

For a complete proof, based on martingale methods applied to the exploration process, see van der Hofstad [41, Sec. 4.5] and the original proof via random walks of Bollobás and Riordan [10].

4.5 Critical Regime and the Critical Window

We now briefly discuss the behaviour of the Erdős–Rényi random graph $G(n,p)$ in the *critical regime*, which occurs when the edge probability lies in a narrow window around the critical value $\lambda = 1$. More precisely, the critical window is defined by

$$p = \frac{1}{n} + \frac{\mu}{n^{4/3}}, \qquad \mu \in \mathbb{R} \text{ fixed.} \tag{4.24}$$

The $n^{-4/3}$ scaling and the $n^{2/3}$ component-size scale, first identified by Bollobás [8] and fully analysed by Łuczak et al. [33] and Aldous [2], arise from the fact that, at criticality, the exploration process behaves like a random walk with vanishing drift and fluctuations of order $n^{1/3}$. As a consequence, the time needed to hit zero is of order $n^{2/3}$, which determines the typical size of the largest components.

In this regime the graph undergoes a transition between the subcritical and supercritical phases, and the component structure exhibits fundamentally new behaviour. The analysis of the critical window is considerably more delicate than in the previous regimes and is treated in detail in Chapter 5 of van der Hofstad [41]. Here we restrict ourselves to an overview of the main results and ideas, focusing on the connection with branching processes.

4.5.1 Breakdown of the classical branching process approximation

In the subcritical and supercritical regimes, the exploration of a connected component can be well approximated by a Galton–Watson branching process with mean offspring $\lambda \neq 1$. At criticality, however, this approximation breaks down.

When $\lambda = 1$, the associated Galton–Watson process is *critical*. As discussed in Chapter 3 (see in particular Theorem 3.3, Theorem 3.6, and Section 3.2.2), such a process becomes extinct almost surely, but its total progeny has infinite expectation and exhibits much larger fluctuations than in the subcritical case [28, 44].

As a consequence, component sizes in the critical Erdős–Rényi graph fluctuate on much larger scales than in the subcritical regime, yet no linear–size giant component has formed.

Moreover, the exploration process can no longer be compared to a branching process with independent increments. Indeed, in the early stages of the exploration, the number of newly discovered neighbours of a vertex is approximately independent and follows a binomial distribution, which can be approximated by a Poisson law (Theorem 3.13). This justifies the Galton–Watson approximation in the subcritical and supercritical regimes.

However, at criticality, the size of the explored component becomes of order $n^{2/3}$, and the pool of unexplored (neutral) vertices is no longer essentially unchanged during the exploration. As a result, the number of available neighbours decreases significantly over time, introducing strong dependencies between successive exploration steps.

In other words, the increments of the exploration process are no longer approximately independent, and the branching process approximation breaks down. A more accurate description requires taking into account the depletion of available vertices, leading to scaling limits described by more complex objects such as the multiplicative coalescent [2, 3, 15, 24].

4.5.2 Size of the largest components

The following theorem summarizes the asymptotic behaviour of the largest connected components in the critical window.

Theorem 4.18 (Critical size of the largest components [41, Ch. 5])**.** *Let p satisfy* (4.24)*. Then, with high probability,*

$$|\mathcal{C}_{\max}| = \Theta(n^{2/3}).$$

In particular, no giant component is present in the critical window.

We do not prove Theorem 4.18 in full; we only give a short heuristic argument indicating why the scale $n^{2/3}$ is the correct one. The full proof requires the scaling theory of the exploration process and can be found in Aldous [2], Janson et al. [24, Ch. 5], and van der Hofstad [41, Ch. 5].

Let (S_t) denote the breadth–first exploration process introduced in Section 4.2. Recall that, in the sparse regime,

$$S_t = S_{t-1} + X_t - 1,$$

where X_t is the number of newly discovered neighbours at step t. Conditionally on the past, $X_t \sim \mathrm{Bin}(N_{t-1}, p)$, where N_{t-1} is the number of neutral vertices; in the early phase $N_{t-1} \approx n$, so that $\mathbb{E}[X_t] \approx np$.

In the critical window (4.24), we have

$$np = n\left(\frac{1}{n} + \frac{\mu}{n^{4/3}}\right) = 1 + \mu n^{-1/3}.$$

Hence the mean increment of the exploration process satisfies $\mathbb{E}[X_t - 1] \approx np - 1 \approx \mu n^{-1/3}$. Summing these mean increments over the first t steps, the deterministic (drift) part of S_t is of order $t\,\mathbb{E}[X_t - 1] \approx t\,n^{-1/3}$. On the other hand, since the increments have bounded variance and are only weakly dependent, the random fluctuations of S_t after t steps are of order $\sqrt{t}$, as for a mean–zero random walk.

The critical scale is obtained when these two contributions are of the same order, that is, when $t\,n^{-1/3} \approx \sqrt{t}$. Squaring both sides yields $t^2 n^{-2/3} \approx t$, and therefore $t \approx n^{2/3}$.

Since the size of a connected component is given by the hitting time of 0 by the exploration process, this shows that the largest critical components appear on the scale $n^{2/3}$. As $n^{2/3} = o(n)$, this scale is much smaller than n, and therefore no giant component appears in the critical window. A rigorous proof, including the convergence of the rescaled component sizes, is given in Aldous [2].

These results provide the correct order of magnitude for the largest components, which will be refined in the scaling description below.

The order $n^{2/3}$ is sharp, and it marks the genuine intermediate scale of the critical window: the largest clusters lie far above the logarithmic components of the subcritical phase, yet they are still negligible compared with n, so that no single component captures a positive fraction of the vertices. This scale was first uncovered by Bollobás [8] and Łuczak et al. [33], while the precise limiting behaviour was obtained by Aldous [2] through continuum random trees and Brownian–motion methods, with self–contained accounts in Janson et al. [24, Ch. 5] and van der Hofstad [41, Ch. 5]. Placing the three regimes side by side, the maximal cluster size jumps from $\Theta(\log n)$ to $\Theta(n^{2/3})$ to $\Theta(n)$ as λ crosses the critical value; the celebrated *double jump*:

$$\Theta(\log n) \;\longrightarrow\; \Theta(n^{2/3}) \;\longrightarrow\; \Theta(n).$$

4.5.3 Scaling limits and exploration processes

A sharper description is obtained by rescaling the exploration walk so that its critical time and space scales become of order one. Since the walk runs for a time of order $n^{2/3}$ and fluctuates on the scale $n^{1/3}$ (as found after Theorem 4.18), the natural normalization is

$$\widetilde{S}_n(u) := n^{-1/3}\, S_{\lfloor un^{2/3} \rfloor}.$$

Aldous [2] identified the limit of $\widetilde{S}_n$ under the critical scaling (4.24): it is a Brownian motion with a parabolic drift, the drift carrying the slightly off–critical bias μ. Connected components are recovered from this limit as the excursions of the reflected process away from zero, and the rescaled sizes of the largest clusters converge jointly to the ordered lengths of these excursions [3, 7]. A single limit theorem thus accounts at once for the emergence of the $n^{2/3}$ scale and for the joint law of the largest critical

components. Developing this theory lies beyond the scope of the present work; we refer the reader to Aldous [2], Janson et al. [24, Ch. 5] and van der Hofstad [41, Ch. 5].

Chapter 5

Conclusions and Further Directions

The aim of this thesis has been to give a self-contained probabilistic account of the phase transition in the Erdős–Rényi random graph $G(n, \lambda/n)$ from the viewpoint of Galton–Watson branching processes. After collecting the basic graph-theoretic background and introducing the model in Chapter 2, we developed in Chapter 3 the theory of Galton–Watson processes needed in the sequel: the extinction-versus-survival dichotomy, the total progeny and its functional equation, the exponential tail bound in the subcritical case, and the binomial and Poisson offspring distributions that arise naturally in the random-graph setting. The core of the analysis was then carried out in Chapter 4, where the breadth-first exploration of a connected component of $G(n, p)$ was bounded above and below by two genuine Galton–Watson processes through the stochastic domination bounds of Theorems 4.2 and 4.4, reducing the analysis of cluster sizes to questions about the total progeny of branching processes.

This comparison gives a sharp picture of the phase transition. In the subcritical regime $\lambda < 1$, the associated branching process becomes extinct almost surely with exponentially decaying total progeny of rate $I_\lambda = \lambda - 1 - \log \lambda$, and this exponential decay translates, via a first- and second-moment argument, into the law of large logarithms $|\mathcal{C}_{\max}|/\log n \to 1/I_\lambda$ in probability, with matching upper and lower bounds (Theorems 4.5 and 4.6). In the supercritical regime $\lambda > 1$, the branching process survives with positive probability ζ_λ, the unique solution in $(0, 1)$ of $\zeta_\lambda = 1 - e^{-\lambda\zeta_\lambda}$; this survival probability turns out to be precisely the asymptotic fraction of vertices in the giant component (Theorem 4.8). The critical regime $\lambda = 1$ has only been touched upon: a heuristic random-walk argument explains the $n^{2/3}$ scale, but a precise description of the joint behaviour of the largest critical components is beyond the scope of this thesis and requires the deep scaling-limit theory of Aldous [2]. These three regimes are summarized in Table 5.1.

Several natural directions of further research lie just beyond the material covered here. A first one concerns the finer geometric structure of the giant component, including its diameter, its typical distances and its local geometry; these have been

Regime	p	Largest component size
Subcritical	$\lambda/n, \quad \lambda < 1$	$\Theta(\log n)$
Critical	$\frac{1}{n} + \frac{\mu}{n^{4/3}}, \quad \mu \in \mathbb{R}$	$\Theta(n^{2/3})$
Supercritical	$\lambda/n, \quad \lambda > 1$	$\Theta(n)$

Table 5.1: The three regimes of the Erdős–Rényi random graph $G(n,p)$ and the corresponding order of magnitude of the largest connected component.

studied in detail by Ding et al. [13] in the strictly supercritical regime and by Lubetzky and Peres [32] in the slightly supercritical regime. A second direction is the extension of the branching-process approach to random-graph models that allow heavy-tailed degree distributions or inhomogeneous edge probabilities, where the Poisson offspring law is replaced by a size-biased degree distribution and the exploration process is compared with multi-type or inhomogeneous branching processes [7, 41]. A third direction is dynamic in nature: viewing p as a time parameter and adding edges one by one turns the phase transition into a genuine dynamical phenomenon, in which the giant component is born through a coagulation mechanism described by the multiplicative coalescent [2, 3]; we refer to Durrett [14, 15] for a systematic exposition.

On a more personal note, working on this thesis has been a very rewarding experience. What attracted me to this topic from the start was the surprising connection between two probabilistic models of completely different origin: the Erdős–Rényi random graph, introduced in the late 1950s as a combinatorial object, and the Galton–Watson process, conceived almost a century earlier to study the extinction of family surnames. The fact that this connection is not just a formal analogy, but the actual key to one of the most studied phase transitions in discrete probability, is what made the topic so interesting to explore. I hope that this thesis can serve as a friendly and self-contained introduction to this beautiful interplay for anyone approaching the subject for the first time.

Appendix Overview

The two appendices that follow collect a few standard probabilistic and mathematical results that are used throughout the thesis but whose proofs or detailed discussion would distract from the main line of argument. Appendix A contains the only piece of probabilistic background that is referenced from the main text without being treated there in full, namely the notion of a discrete-time Markov chain, which is needed to describe the Galton–Watson process of Chapter 3 as a Markov chain on $\mathbb{N}_0$. Appendix B gathers the mathematical tools that recur in the proofs: the basic probabilistic inequalities (Markov, Chebyshev, second moment method, and the Chernoff–Cramér bound), the notion of stochastic domination together with the coupling method, and the O, o, Θ, Ω asymptotic notation.

Appendix A

Probabilistic Background

Definition A.1 (Markov chain)**.** A sequence of random variables $(X_t)_{t\geq 0}$ with values in a countable set E is a *Markov chain* if, for every $t \geq 0$ and every $x_0, \dots, x_t, y \in E$,

$$\mathbb{P}(X_{t+1} = y \mid X_t = x_t,\, X_{t-1} = x_{t-1}, \dots, X_0 = x_0) \;=\; \mathbb{P}(X_{t+1} = y \mid X_t = x_t).$$

The chain is called *time-homogeneous* if the right-hand side does not depend on t, in which case the function $p(x, y) := \mathbb{P}(X_{t+1} = y \mid X_t = x)$ is called the *transition kernel* of the chain and satisfies $\sum_{y \in E} p(x, y) = 1$ for every $x \in E$.

Remark A.2 (Galton–Watson processes are Markov chains)**.** The Galton–Watson branching process $(Z_t)_{t\geq 0}$ introduced in Chapter 3 is a time-homogeneous Markov chain on $\mathbb{N}_0$. Indeed, given $Z_t = k$, the next generation has the explicit representation $Z_{t+1} = \sum_{i=1}^{k} X_{t,i}$, where the $X_{t,i}$ are i.i.d. copies of the offspring variable X that are independent of $Z_0, \dots, Z_t$; hence the conditional distribution of Z_{t+1} given the whole past depends only on Z_t.

Appendix B

Mathematical Tools

Basic probabilistic inequalities

Remark B.1. Markov's inequality controls the probability that a non-negative random variable is large in terms of its expectation, and is the basis of all *first-moment arguments*. In this thesis it is used in Chapter 4, together with the exponential tail bound of Proposition 3.7, to derive the upper bound on the size of the largest subcritical component (Theorem 4.5).

Proposition B.2 (Chebyshev's inequality)**.** *Let X be a random variable with finite variance. Then for every $a > 0$,*

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

Remark B.3. Chebyshev's inequality follows from Markov's inequality applied to the non-negative random variable $(X - \mathbb{E}[X])^2$. In this thesis it is used in Chapter 4 to prove the concentration of the random variable $Z_{\geq k_n}$ around its mean and thereby establish the law of large numbers for the giant component (Theorem 4.8).

Proposition B.4 (Second moment method)**.** *Let X be a non-negative random variable with $\mathbb{E}[X^2] < \infty$. Then*

$$\mathbb{P}(X > 0) \;\geq\; \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

Remark B.5. The second moment method is the standard tool to show that a non-negative random variable is positive with high probability when its variance is small relative to the square of its expectation [4, Ch. 4]. In this thesis it is applied to $X = Z_{\geq k}$ in the proof of Theorem 4.6, to show that with high probability at least one vertex initiates a connected component of size of order $\log n / I_\lambda$.

Remark B.6 (Chernoff–Cramér bound)**.** A standard application of Markov's inequality to the non-negative random variable $e^{\theta X}$ gives, for every $\theta > 0$ and $a \in \mathbb{R}$,

$$\mathbb{P}(X \geq a) \;\leq\; e^{-\theta a}\, \mathbb{E}[e^{\theta X}].$$

Optimising over $\theta > 0$ yields the *Chernoff–Cramér bound* $\mathbb{P}(X \geq a) \leq \exp(-I(a))$, where $I(a) = \sup_{\theta \geq 0}(\theta a - \log \mathbb{E}[e^{\theta X}])$ is the Legendre transform of the logarithmic moment generating function of X; see Dembo and Zeitouni [11, Ch. 2]. This is the technique used in the proof of Proposition 3.7 to obtain the exponential tail of the total progeny of a subcritical Poisson Galton–Watson process.

Stochastic domination and coupling

Definition B.7 (Stochastic domination)**.** For two random variables X and Y taking values in $\mathbb{R}$, we say that X is *stochastically dominated* by Y, and we write $X \preceq_{\mathrm{st}} Y$, if

$$\mathbb{P}(X > t) \;\leq\; \mathbb{P}(Y > t) \qquad \text{for every } t \in \mathbb{R}.$$

Definition B.8 (Coupling)**.** A *coupling* of two random variables X and Y with given marginal laws is a pair of random variables $(\widetilde{X}, \widetilde{Y})$ defined on a common probability space, such that $\widetilde{X}$ has the same distribution as X and $\widetilde{Y}$ has the same distribution as Y. If a coupling can be constructed in which $\widetilde{X} \leq \widetilde{Y}$ almost surely, then $X \preceq_{\mathrm{st}} Y$.

The converse is also true: $X \preceq_{\mathrm{st}} Y$ if and only if such a monotone coupling exists; see Shaked and Shanthikumar [38, Ch. 1] and Lindvall [31]. The coupling method is the technical backbone of the stochastic domination bounds developed in Section 4.2.

Example B.9 (Binomial comparison)**.** Let $0 < m \leq n$ be integers and $p \in [0, 1]$, and let $B_1, \ldots, B_n$ be i.i.d. Bernoulli random variables with parameter p. Setting $\widetilde{X} := \sum_{i=1}^{m} B_i$ and $\widetilde{Y} := \sum_{i=1}^{n} B_i$ produces a coupling of $\mathrm{Bin}(m, p)$ and $\mathrm{Bin}(n, p)$ in which $\widetilde{X} \leq \widetilde{Y}$ almost surely; by Definition B.8 this gives $\mathrm{Bin}(m, p) \preceq_{\mathrm{st}} \mathrm{Bin}(n, p)$. This is the elementary fact used in the proof of Theorem 4.4 to compare the exploration process of $G(n, p)$ with a Galton–Watson process of binomial offspring distribution with reduced parameter.

List of symbols

Let $n \in \mathbb{N}$, $p \in [0,1]$, $\lambda > 0$. Let $G = (V, E)$ be a simple graph and $(Z_t)_{t \geq 0}$ a Galton–Watson process with offspring variable X.

Asymptotic notation

$f = O(g)$	$\exists\, c, n_0 > 0$ such that $\lvert f(n)\rvert \leq c\, g(n)$ for all $n \geq n_0$
$f = o(g)$	$f(n)/g(n) \to 0$ as $n \to \infty$
$f = \Theta(g)$	$f = O(g)$ and $g = O(f)$
$f = \Omega(g)$	$\exists\, c, n_0 > 0$ such that $\lvert f(n)\rvert \geq c\, g(n)$ for all $n \geq n_0$

Probability

$\mathbb{P}$, $\mathbb{E}$, Var	Probability, expectation, variance
$\mathbb{P}_{n,p}$	Law of $G(n,p)$
$X \xrightarrow{d} Y$	Convergence in distribution
$X \xrightarrow{\mathbb{P}} Y$	Convergence in probability
$\mathrm{Bin}(n,p)$	Binomial distribution with parameters n and p
$\mathrm{Pois}(\lambda)$	Poisson distribution with mean λ
$\mathcal{N}(0,\sigma^2)$	Normal distribution with mean 0 and variance σ^2

Graph theory

$G = (V, E)$	Simple undirected graph
K_n	Complete graph on n vertices
$[n] = \{1, \ldots, n\}$	Standard vertex set
d_v	Degree of vertex $v \in V$
$G[V_H]$	Induced subgraph of G on $V_H \subseteq V$
$\mathrm{dist}_G(u,v)$	Graph distance between u and v
$\mathcal{C}(v)$	Connected component of vertex v
$\mathcal{C}_{\max}$	Largest connected component and its size
$\mathcal{C}_{(2)}$	Second-largest connected component

Erdős–Rényi random graph

$G(n,p)$	Random graph on $[n]$; each edge included independently with prob. p
$p = \lambda/n$	Sparse regime, $\lambda > 0$ fixed

$I_\lambda = \lambda - 1 - \log \lambda$	Large-deviation rate for Pois(λ) at $x = 1$
$Z_{\geq k}$	Vertices belonging to components of size $\geq k$
Breadth-first exploration	
S_i	Number of active vertices after i steps
N_i	Number of neutral vertices after step i
X_i	Newly discovered vertices at step i
$T = \min\{i \geq 1 : S_i = 0\}$	Stopping time; equals $\lvert\mathcal{C}(v)\rvert$
$T^{\geq}$, $T^{\leq}$	Total progeny of upper / lower bounding process
Galton–Watson branching process	
$\mathrm{GW}(X)$	Law of the process with offspring X
Z_t	Population size in generation t
$\mu = \mathbb{E}[X]$	Mean offspring number
$G_X(s) = \mathbb{E}[s^X]$	Probability generating function (PGF) of X
η	Extinction probability; smallest fixed point of G_X in $[0, 1]$
η_λ	Extinction prob. for Pois(λ); smallest root of $\eta = e^{\lambda(\eta-1)}$
$\zeta_\lambda = 1 - \eta_\lambda$	Survival probability for Pois(λ) offspring
$T = \sum_{n\geq 0} Z_n$	Total progeny
$G_T(s) = \mathbb{E}[s^T]$	PGF of total progeny; satisfies $G_T(s) = s\, G_X(G_T(s))$
$\mu_\lambda \in (0, 1)$	Conjugate parameter; $\mu_\lambda e^{-\mu_\lambda} = \lambda e^{-\lambda}$
σ_λ^2	Limiting variance in the CLT for $\lvert\mathcal{C}_{\max}\rvert/\sqrt{n}$
Critical window	
$p = \frac{1}{n} + \mu\, n^{-4/3}$	Edge probability in the critical window, $\mu \in \mathbb{R}$ fixed
$\widetilde{S}_n(u) = n^{-1/3} S_{\lfloor u\, n^{2/3} \rfloor}$	Rescaled exploration process
$I(x) = \sup_{\theta\geq 0}(\theta x - \Lambda(\theta))$	Legendre transform; large-deviation rate function

Bibliography

[1] Miklós Ajtai, János Komlós, and Endre Szemerédi. Largest random component of a k-cube. *Combinatorica*, 2(1), 1982.

[2] David Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Annals of Probability*, 25(2), 1997.

[3] David Aldous and Vlada Limic. The entrance boundary of the multiplicative coalescent. *Electronic Journal of Probability*, 3(3), 1998.

[4] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley, 4 edition, 2016.

[5] K. B. Athreya and P. E Ney. *Branching Processes*. Springer, 1972.

[6] Andrew D. Barbour, Lars Holst, and Svante Janson. *Poisson Approximation*. Oxford University Press, 1992.

[7] Shankar Bhamidi, Remco van der Hofstad, and Johan S. H. van Leeuwaarden. Scaling limits for critical inhomogeneous random graphs. *The Annals of Probability*, 2010.

[8] Béla Bollobás. The evolution of random graphs. *Transactions of the American Mathematical Society*, 286(1), 1984.

[9] Béla Bollobás. *Random Graphs*. Cambridge University Press, 2 edition, 2001.

[10] Béla Bollobás and Oliver Riordan. Asymptotic normality of the size of the giant component via a random walk. *Journal of Combinatorial Theory, Series B*, 102 (1), 2012.

[11] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*. Springer, 2 edition, 1998.

[12] Reinhard Diestel. *Graph Theory*. Springer, 2 edition, 2000.

[13] Jian Ding, Eyal Lubetzky, and Yuval Peres. Anatomy of the giant component: the strictly supercritical regime. *European Journal of Combinatorics*, 35, 2014.

[14] Rick Durrett. *Random Graph Dynamics*. Cambridge University Press, 2007.

[15] Rick Durrett. *Dynamics on Graphs*. Cambridge University Press, 2024.

[16] Paul Erdős and Alfréd Rényi. On random graphs I. *Publicationes Mathematicae Debrecen*, 6, 1959.

[17] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 5, 1960.

[18] William Feller. *An Introduction to Probability Theory and Its Applications, Vol. 1*. Wiley, 3 edition, 1968.

[19] Alan Frieze and Michał Karoński. *Introduction to Random Graphs*. Cambridge University Press, 2016.

[20] Edgar N. Gilbert. Random graphs. *Annals of Mathematical Statistics*, 30(4), 1959.

[21] Patsy Haccou, Peter Jagers, and Vladimir A. Vatutin. *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press, 2005.

[22] T. E. Harris. *The Theory of Branching Processes*. Springer, 1963.

[23] Peter Jagers. *Branching Processes with Biological Applications*. Wiley, 1975.

[24] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random Graphs*. Wiley, 2000.

[25] Richard M. Karp. The transitive closure of a random digraph. *Random Structures & Algorithms*, 1(1), 1990.

[26] Harry Kesten and Bernt P. Stigum. A limit theorem for multidimensional Galton–Watson processes. *Annals of Mathematical Statistics*, 37(5), 1966.

[27] Marek Kimmel and David E. Axelrod. *Branching Processes in Biology*. Springer, 2 edition, 2015.

[28] Andrey N. Kolmogorov. Zur Lösung einer biologischen Aufgabe. *Communications of the Mathematical Society of Kharkov*, 2(13), 1938.

[29] Michael Krivelevich and Benny Sudakov. The phase transition in random graphs: a simple proof. *Random Structures & Algorithms*, 43(2), 2013.

[30] Lucien Le Cam. An approximation theorem for the Poisson binomial distribution. *Pacific Journal of Mathematics*, 10(4), 1960.

[31] Torgny Lindvall. *Lectures on the Coupling Method*. Dover Publications, 2002.

[32] Eyal Lubetzky and Yuval Peres. Anatomy of a young giant component in the random graph. *Random Structures & Algorithms*, 42(2), 2013.

[33] Tomasz Łuczak, Boris Pittel, and John C. Wierman. The structure of a random graph at the point of the phase transition. *Transactions of the American Mathematical Society*, 341(2), 1994.

[34] Russell Lyons and Yuval Peres. *Probability on Trees and Networks*. Cambridge University Press, 2016.

[35] Russell Lyons, Robin Pemantle, and Yuval Peres. Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Annals of Probability*, 23(3), 1995.

[36] Anders Martin-Löf. *Random Graphs: An Introduction*. Lecture Notes, Stockholm University, 2014.

[37] Boris Pittel. On tree census and the giant component in sparse random graphs. *Random Structures & Algorithms*, 1(3), 1990.

[38] Moshe Shaked and J. George Shanthikumar. *Stochastic Orders*. Springer, 2007.

[39] Joel Spencer. The giant component: The phase transition in random graphs. *Notices of the AMS*, 41(5), 1994.

[40] V. E. Stepanov. On the probability of connectedness of a random graph $G_m(t)$. *Theory of Probability and its Applications*, 15(1), 1970.

[41] Remco van der Hofstad. *Random Graphs and Complex Networks, Vol. 1*. Cambridge University Press, 2017.

[42] Henry William Watson and Francis Galton. On the probability of the extinction of families. *Journal of the Anthropological Institute of Great Britain and Ireland*, 4, 1875.

[43] Douglas B. West. *Introduction to Graph Theory*. Prentice Hall, 2 edition, 2001.

[44] Akiva M. Yaglom. Certain limit theorems of the theory of branching random processes. *Doklady Akademii Nauk SSSR*, 56, 1947.

Declaration on the Use of Generative AI

I used OpenAI ChatGPT and Anthropic Claude as auxiliary tools in the writing process to improve the clarity, readability, and linguistic correctness of selected passages, as well as the structure and readability of the write-up of some proofs. All suggestions were reviewed and adapted by the author, who retains full responsibility for the accuracy and integrity of this thesis.