



From Erdos-Renyi to Achlioptas: The Birth of a Giant

Link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:38811522>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

Contents

1	Introduction	3
1.1	Introduction	3
1.2	Outline of the Paper	4
1.3	Why should I read this paper?	5
2	The Tree-Counting Technique: Phase Transition in Erdős-Rényi Graphs	6
2.1	Brief Outline of Results	6
2.1.1	Why $t^* = n/2$? A taste of branching process estimations	6
2.1.2	Preview of the remaining section	7
2.2	Combinatorial Preliminaries: Tree Counting	7
2.3	Pre-transition: Sparse Forests	10
2.4	Phase Transition: the Giant Component	11
2.4.1	Step 1: Proving the Component Gap Theorem	12
2.4.2	Step 2: Estimating the vertices in large components	14
2.4.3	Step 3: Erdős-Rényi Sprinkling Argument: Creation of the giant component	15
2.5	Post-Phase Transition: the Growth of the Giant Component	16
2.6	Conclusion and Renormalization	17
2.6.1	Brief Recap and Renormalization	17
2.6.2	Weakly supercritical phase	18
3	Generalization of Erdős-Rényi Random Graph Processes: a Brief Overview and Introduction to Achlioptas Processes	19
3.1	Modelling Real World Networks	19
3.1.1	Clustering	20
3.1.2	Power-Law	20
3.2	How universal is the phase transition behavior of Erdős-Rényi Processes? Introduction to Achlioptas Processes	21
3.2.1	Achlioptas Processes: A Formal Definition	21
3.2.2	Bounded-Achlioptas Processes: The “Universality” of Erdős-Rényi Phase Transition	22
3.2.3	Key Techniques	23
4	Technique 1: Branching Processes	24
4.1	Single-Variable Branching Processes: Criticality at $\mu = 1$	24
4.1.1	Basic Definition and Results	25
4.1.2	Subcriticality: Bounds on Total Population	25
4.1.3	Bounds on Survival Probability	26
4.2	Sesqui-type Processes and Branching Process Families	28
4.2.1	Basic Definition and Results	28
4.2.2	Branching Process Families	28

5	Technique 2: DEM	30
5.1	Toy Examples	30
5.1.1	Balls and Bins	30
5.1.2	Independence number of $G_{n,r}$	31
5.2	Wormald's Theorem	32
5.3	Applications to Achlioptas Processes: Small Components and Susceptibility	33
5.3.1	Local Convergence	33
5.3.2	Susceptibility	35
6	Bounded-Size Achlioptas Processes: a Universality Class for Erdős-Rényi Phase Transition	36
6.1	Statement of the Main Result	36
6.1.1	Simplification of the Result	38
6.2	Preliminaries: Two-round Exposure, The Poissonized Graph, and the Concentration of the Parameter List	38
6.2.1	Two-round Exposure, the Auxillary Graph H_i , and the parameter list $\mathcal{G}_i$	38
6.2.2	Evolution of $\mathcal{G}_i$: Applications of DEM	41
6.3	Component size distribution: coupling arguments	45
6.3.1	Neighborhood Exploration Process: From component size distribution N_j to the Associated Random Walk T	45
6.3.2	The Branching Process: from T to $\mathcal{S}$	47
6.3.3	Analysis of the Branching Process $\mathcal{S}$	50
6.4	Putting Things Together: The Final Proof	51
6.4.1	Small Components	51
6.4.2	Size of the largest component	52
6.4.3	Size of the largest component: Supercritical Case	52
6.5	Summary and Takeaway	53
6.6	Conclusion and Future Work	54
	APPENDICES	56
A	Proof of Theorem 1 and other Combinatorial Results	57
B	Other Deferred Proofs from the Erdős-Rényi Section	61
B.1	Reduction from $G(n, t)$ to $G(n, p)$	61
B.2	Proof of the Connectivity Result	61
B.3	Plot of ρ_{ER}	62
B.4	An alternative approach for the subcritical Erdős-Rényi graph: the random walk approach . .	62
C	Sketch of the Proof for Sesqui-type Branching Process	64
D	Proof of Relevant Inequalities: Hoeffding-Azuma, McDiarmid, and Chernoff	66
E	Proof that $\mathcal{S}_t$ is a t_c-critical branching process family	68

Chapter 1

Introduction

“With $t = -10^6$, say we have feudalism. Many components (castles) are each vying to be the largest. As t increases, the components increase in size and a few large components (nations) emerge. An already large France has much better chances of becoming larger than a smaller Andorra. The largest components tend to merge and by $t = 10^6$ it is very likely that a giant component, the Roman Empire, has emerged. With high probability this component is nevermore challenged for supremacy but continues absorbing smaller components until full connectivity - One World - is achieved.”

Noga Alon, Joel H. Spencer, The Probabilistic Method

1.1 Introduction

Ever since the seminal work of Erdős and Rényi [14], the study of random graphs has garnered interdisciplinary attention. Social scientists ([31]) have used random graph processes to model real-life networks, as well as to detect latent communities. Perhaps closer to the original intent of Erdős and Rényi, mathematicians and computer scientists have used random graph theory extensively to prove many results in graph theory and combinatorics via the probabilistic method. Furthermore, random graph processes have also been increasingly studied as an independent mathematical object.

A common and robust qualitative feature of random graphs is the surprisingly early and sharp emergence of salient global features. Let us take connectivity as an example. Obviously, one needs at least $n - 1$ edges to connect a graph of order n . However, one of the striking original results by Erdős and Rényi states that one does not need “much” more than that – adding $O(n \log n)$ random edges will ensure that the graph is connected with probability $1 - o(1)$! Furthermore, the bound $n \log n$ is sharp: there is a constant c such that adding fewer than $cn \log n$ edges will imply that the graph is not connected with high probability. Such a sharp qualitative change in the global random graph when one crosses a critical threshold is defined to be a phase transition in the random graph.

The existence of a phase transition in a random graph is at first glance striking and counterintuitive – given that the graph is random, how can it seem to have “almost sure” properties? Secondly, is it natural to have a jarring, qualitative change when one continuously varies a parameter? However, classic models in probability theory often exhibit such features. For example, Kolmogorov’s 0-1 law demonstrates that properties that depend only on the tail σ -algebra, such as the convergence of a series, must have probability 0 or 1. Classic examples include the law of large number results. As for phase transition, a well-known and relevant example is the almost sure extinction of a Galton-Watson branching process of expected offspring less than 1, and the positive survival probability for offspring mean greater than 1. Regardless, the global, almost-sure properties and the phase transition behavior in random graphs have received much publicity and have been studied extensively.

While we started with the example of the connectivity phase transition, the focus of this expository paper will be another, closely related type of phase transition, which was also first discovered by Erdős and Rényi. The following table shows the size of the largest component of $G(10000, p)$, where each edge is independently added with probability p .

Erdős Rényi Giant Component					
p	L_1	p	L_1	p	L_1
0.7/n	38	1/n	902	1.3/n	4353
0.8/n	54	1.1/n	2267		
0.9/n	325	1.2/n	2888		

Note that around $p = 1/n$, there is a sharp increase in the number of components belonging to the largest component. Roughly speaking, whereas the largest component size is $o(n)$ (or more precisely, $O(\log n)$) for smaller values of p , when p grows larger than $1/n$, suddenly the size of the largest component grows to $\Theta(n)$. The giant component has emerged.

Roughly speaking, the giant component phase transition is qualitatively similar to the connectivity phase transition: as the graph grows increasingly connected, so do its connected components. Once a giant component emerges, it is dominant: no other component will ever grow larger, and will eventually be absorbed into the giant component. I was drawn to this particular phase transition, as I believed this to be qualitatively the more robust phenomenon: a single isolated vertex will imply that a graph is no longer connected, while the existence of a component of “non-negligible” size is robust to the addition/deletion of edges and vertices.

1.2 Outline of the Paper

In this expository paper, I thus aim to give an exposition on the emergence of the giant component in random graph processes. Among the myriad random graph models, I shall start with the classic Erdős-Rényi model, where either m out of the $\binom{n}{2}$ edges are selected uniformly at random ($G_{n,m}$), or each edge added independently with probability p ($G_{n,p}$). Then, we shall work towards an analysis of the Achlioptas processes, which is a generalization of the original Erdős-Rényi model.

As the Erdős-Rényi model forms the baseline canonical model for all random graph models, the analysis of the giant component phase transition in the Erdős-Rényi model merits special importance. In Chapter 2, I shall closely follow the exposition of Bollobás ([9]). While there are much faster proofs of the phase transition itself (see, for example, [17]), Bollobás’ proof is valuable to study for the following reasons.

1. The key technique (tree-counting) is completely elementary.
2. We obtain far more granular data, such as the relative frequencies of the components of a given size, the average complexity of a given component (whether they are trees, unicyclic, or more complex), the speed of the growth of the giant component, and so forth. In particular, we find that save for the giant component, the other components are all sparse trees, which get swallowed up by the giant component.
3. Techniques aside, the basic idea of the proof, which is to prove the accumulation of enough points in “large components”, and then conclude that they coalesce quickly into the giant component, will be similar in spirit to the more involved proof we will give regarding more general random graph processes.

After demonstrating the emergence of the giant component for Erdős-Rényi graphs, I shall take a brief detour in Chapter 3, to give a short overview of some of the extensions to the Erdős Rényi model. The generalizations can be broadly categorized into two types: either they try to better fit certain salient features of real-life networks, or they attempt to provide mathematical generalizations of the original model. One such example of the latter case that has recently received a lot of attention is the Achlioptas process.

At its simplest definition, the Achlioptas process selects two edges at random, and select an edge among the two according to a rule depending only on the component sizes of the endpoints of each edges. Inspired

by the “power of two choices” paradigm([20]), Achlioptas hoped to demonstrate that the accumulation of these small decisions could result in significant changes in the qualitative behavior of the random graph.

However, despite the generality of the edge-selection rule, as long as the rule is bounded-size, namely, it treats all components larger than K the same, the random graph process exhibits a similar phase transition behavior to the Erdős-Rényi model. Hence, the bounded-size Achlioptas processes form a greater universality class of Erdős-Rényi graphs with respect to its phase-transition behavior.

The remainder of this paper seeks to give a proof of this striking universality result. While this has been shown originally in Spencer and Wormald 2007 ([29]), I shall focus on the exposition given by Riordan and Warnke 2017 ([25]), which obtains more precise asymptotics for component size distributions, as well as stronger notions of convergence. Both papers are similar in spirit, relying on the following three ingredients:

1. The exploration process on the original random graph
2. A coupled idealized branching process
3. The differential equation method governing the evolution of normalized random graph quantities

This “trinity” is essential not only in analyzing Achlioptas processes, but general non-Erdős-Rényi random graph processes, where there is significant dependence between past and future edges. To make the exposition as self-contained as possible and give a taste of these important tools, I shall spend some time giving a sketch of the main results on branching processes and the differential equation method in Chapters 4 and 5.

In Chapter 6, I shall bring the results developed in Chapters 4 and 5 to finally give the proof of the universal phase transition in Achlioptas processes. For the ease of exposition, we shall prove weaker results, such as proving pointwise instead of uniform convergence of key quantities. Hopefully, the proof will showcase the interplay between the three key ingredients (the exploration process, the branching process, and the differential equation method), without going into the exhausting detail of the original proof (which is almost 100 pages long). After we prove the theorem, we shall summarize and conclude, illustrating some open questions as well as relevant results not covered in this exposition.

1.3 Why should I read this paper?

- *For mathematicians:* The phase transition result for bounded-size Achlioptas processes is an example of a universality result, which states that certain qualitative features remain robust to specification details. Regardless of the exact choice of the edge selection rule, the accumulation of random edges for bounded-size Achlioptas process results in a universal Erdős-Rényi-like phase transition behavior. Study of such universality results attract wide interest in mathematics, from probability to even number theory (see, for example, [21]’s conjecture on the spacing of the zeros of the ζ function). Other examples include Tao et al ([32]), which shows “universal” convergence of the empirical spectral distributions to the semi-circle law, regardless of the specification of the distribution of the matrix entries, or more classically, the central limit theorem and the law of large numbers.
- *For computer scientists:* While the theory of Achlioptas processes itself is probably of less inherent interest to computer scientists, the proof of the phase transition result for Achlioptas processes showcases key techniques, both the differential equation method and the branching process theory, developed to analyze general discrete processes. Along with the proof of the main result, I have provided separate expositions of these techniques, which will be of independent importance in the analysis of discrete processes and randomized algorithms.
- *For students:* In many ways, this expository paper represents my own journey to learn this beautiful and fascinating result. Consequently, I have aimed to make the exposition as self-contained as possible, only assuming a background in basic probability. The probability techniques used in the proofs, such as martingale arguments, Chebyshev and Chernoff bounds, and Poisson process analysis, have wide applicability in proofs in probability and statistics. By working through the exposition, I hope students of various mathematical training can pick up these techniques to use for future problem-solving.

Chapter 2

The Tree-Counting Technique: Phase Transition in Erdős-Rényi Graphs

2.1 Brief Outline of Results

In this section, I shall give an exposition of the classic results concerning the emergence of the giant component in Erdős Rényi random graph processes. While the results involving the emergence of a giant component originally appeared in the seminal work of [14], I shall closely follow the path taken by [9], which provides a more detailed understanding of the exact phase transition.

Denote G_t as the Erdős-Rényi random graph process, obtained by adding at each step an edge selected uniformly at random among the remaining edges. The basic results concerning the emergence of a giant component can be summarized as follows:

1. Prior to the critical time $t = n/2$, (also denoted as the percolation threshold) the largest component has size at most $\log n$, and most components are sparse, i.e. trees.
2. The time $t = n/2$, corresponding to $p = 1/n$, is critical in the following sense: shortly after $n/2$, there is a unique component of order at least $n^{2/3}$, and all other components have fewer than $n^{2/3}/2$ vertices.
3. After the critical time $t = n/2$, the giant component grows rather regularly, roughly at 4 times as fast as t , and the unique largest component has size ϵn after the critical time.

The referenced [9], a slightly more streamlined version of the original paper [7], mainly uses tree-counting and combinatorial bounds. While this technique does not generalize to models where there is interdependence between the edges, it is able to give fairly sharp results for the Erdős-Rényi model. Moreover, the combinatorial techniques involved are, albeit tedious, quite elementary. Furthermore, the rich qualitative results regarding the phase transition turn out to hold robustly for more general random graph processes. For its inherent interest and to motivate future sections, I shall describe and prove some of the major results for Erdős-Rényi graphs, devoting time and space to hammer out some of the details in the proof.

2.1.1 Why $t^* = n/2$? A taste of branching process estimations

Before we begin the exposition of the proof, here is a quick intuition as to why $t = n/2$, or equivalently, $p = 1/n$, is the critical threshold for the phase transition behavior. We shall sketch the classic arguments relying on branching processes and random walks, which give a clear hint why $t = n/2$ is the percolation time, or equivalently $p = t/(n^2/2) = 1/n$.

While these techniques will not be used in the exposition of this section (only tree-counting and basic inequalities will be used), they are not only more intuitive, but also will become relevant for general random graph processes.

The idea itself is simple: instead of viewing the random graph at once, one can pick a random vertex and explore its neighbors. Thus, the neighborhood exploration process can be well-approximated by starting from a vertex, and regarding its neighbors as offsprings of a branching process, with an offspring distribution of $\text{Bin}(n-1, p)$. Of course, the approximation is not perfect: the newly introduced neighbors of the branching process may already be explored, or perhaps overlap. This corresponds to the component having cycles.

If the component is small, or more specifically, the number of particles in the branching process is dominated by n , such events happen with vanishingly small probability. However, when one seeks to investigate the giant component with this methodology, such assumptions may not hold, and hence one must be careful to check that the approximation is good.

Regardless, if the branching process approximation is good, note that the expectation of the offspring is roughly $\mu = np$. It is well-known that μ governs the long-run behavior of the branching process: it can be shown (and will be shown when we return to branching processes in detail) that for $\mu \leq 1$, the process almost surely terminates, while for $\mu > 1$, the process never terminates with positive probability.

Taking somewhat of a leap of faith, one can see that a terminating branching process implies a small component, while a non-terminating process corresponds to a component of size $\omega(n)$, i.e. growing to ∞ as $n \mapsto \infty$. In fact, as [13] demonstrates, for $\mu > 1$, $P(\lim X_t/\mu^t \neq 0) > 0$, so X_t intuitively grows quite large with positive probability.

In fact, by taking another Herculean leap of faith, one can postulate that the probability that the branching process does not terminate corresponds to the fraction of vertices belonging to the giant component of G , as this corresponds to the probability of selecting a vertex that belongs to a “large” component. This turns out to be true quite robustly. Bollobás [8] gives a heuristic explanation for Erdős-Rényi, which can be verified via the explicit formula for ρ shown later.

As always, it is helpful to see the same argument made in a different way, this time using the random walk associated with the branching process. Given B a branching process, one can consider the following neighborhood exploration process:

1. Initialize $S = \{v_0\}$, the first generation. Iterate the following until $S = \emptyset$:
2. Pick an element $v \in S$ and add to S all of its offsprings.
3. Remove v from S .

Then, the random walk associated with the branching process is simply the size of S . Let X_n be the change in $|S|$, which is distributed as $X_n \sim -1 + \text{Bin}(n, \lambda/n)$. Then, note $S_n - (\lambda - 1)n$ is a martingale, and define T to be the stopping time when S_n hits 0. By classic martingale techniques, one can use Wald’s equation to conclude $E[S_T - S_0] = (\lambda - 1)E[T]$, and hence $E[T] = 1/(1 - \lambda)$. Hence, for $\lambda < 1$, we obtain a finite estimation for expected component size, while $\lambda \geq 1$ implies a potentially infinite component size.

2.1.2 Preview of the remaining section

We shall organize the proof of the emergence of the giant component in “chronological” order. First, we shall state a key combinatorial result, which contain most of the tree-counting technique needed for this section. Intuitively, we show rigorously that “most” components of a fixed size are trees.

After the preliminaries, we proceed in chronological order. We first describe the structure of the random graph before the phase transition, where most of the vertices belong to trees of small order. Next, we shall dive into the heart of the proof, the emergence of the giant component. Finally, we shall give a brief description of the growth of the giant component and the structure of the random graph post-phase transition.

2.2 Combinatorial Preliminaries: Tree Counting

Let $E_t(k)$ be the expected number of components of order k in G_t . To evaluate $E_t(k)$, denote $X(k, d)$ as the number of components of order k and size $k + d$. Let $N = \binom{n}{2}$ be the total number of possible edges.

By selecting the k vertices to be part of the (k, d) components and ensuring that the remaining $t - k - d$ edges connect the remaining $n - k$ vertices with each other, one obtains:

$$E_t[X(k, d)] = \binom{n}{k} C(k, d) \binom{\binom{n-k}{2}}{t-k-d} / \binom{N}{t} \quad (2.1)$$

where $C(k, d)$ is the number of distinct connected components of order k and size $k + d$. For example, $C(k, -1)$ is the number of distinct trees with labels 1 through k .

Almost equivalently, one can also define $E_p[X(k, d)]$ for $G(n, p)$, with:

$$E_p[X(k, d)] = \binom{n}{k} C(k, d) p^{k+d} (1-p)^{kn-k(k+3)/2-d} \quad (2.2)$$

The first critical result we shall need is that the expected number of components of size k , i.e. $\sum_{d \geq -1} E_p(k, d)$ is of the same order as the number of trees of size k , i.e. $E_p(k, -1)$. We shall demonstrate this by proving the following upper bound for $C(k, d)$.

Theorem 1. *There is an absolute constant c such that $C(k, d) \leq cd^{-d/2} k^{k+(3d-1)/2}$.*

We shall defer the proof of this theorem to the appendix. Along with this upper bound, we shall state more precise results for $d = -1$ and 0, corresponding to trees and unicyclic components. We shall also defer the proofs to the appendix.

Theorem 2. 1. *Cayley's Theorem: the number of tree components of size k is $C(k, -1) = k^{k-2}$.*

2. *The number of unicyclic components is:*

$$C(k, 0) = \frac{1}{2} k^{k-1} \sum_{r=3}^k \prod_{j=1}^{r-1} \left(1 - \frac{j}{k}\right) = \frac{1}{2} (k-1)! \sum_{j=0}^{k-3} \frac{k^j}{j!} \sim (\pi/8)^{1/2} k^{k-1/2} \quad (2.3)$$

One can obtain a quick corollary: fixing k , most of the k -components will be trees, as long as $p = O(1/n)$.

Corollary 1 (Trees dominate in expectation). *Let $E_p(k)$ be the expected number of k -components in $G(n, p)$, and $E_p^T(k)$ be the expected number of trees of size k . Assume $p = O(1/n)$, and $k < n^{2/3}$. Then, there is an absolute constant C such that $E_p(k) \leq C \cdot E_p^T(k)$. In general, for a given complexity $d \geq -1$, the expected number of (k, d) components for all $d \geq \bar{d}$ is bounded above by a constant factor by the expected number of $(k, \bar{d})$ components.*

Proof. We simply compute the expectation directly, and use the combinatorial bound to bound the sum over d by a geometric series. As a consequence of our upper bound, one has $C(k, d) \leq c_0 6^{-d} k^{k+(3d-1)/2}$. By our assumption, let us set $p = \frac{D}{n}$, $D < 2$ (for convenience). Plugging this in, we obtain:

$$\begin{aligned} E_p(k) &\leq \binom{n}{k} c_0 \sum_{d=-1}^{k(k-1)/2-k} 6^{-d} k^{k+(3d-1)/2} p^{k+d} (1-p)^{kn-k(k+3)/2-d} \\ &\leq \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{kn-k(k+3)/2+1} \cdot c_0 \sum_{d=0}^{\infty} \left(\frac{p}{6(1-p)} k^{3/2} \right)^d \leq E_p^T(k) \cdot c_0 \sum_{d=0}^{\infty} \left(\frac{1}{3} \right)^d \end{aligned} \quad (2.4)$$

which gives us the bound, as desired. The same logic gives us the statement for more general $\bar{d}$. □

The following asymptotic formula for $E_p^T(k)$ will be useful:

Corollary 2 (Tree Asymptotics). *Let $p = c/n$. Then, one obtains:*

$$\begin{aligned} E_p^T(k) &= \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{kn-k(k+3)/2+1} \sim \frac{1}{c} n \prod_{i=0}^{k-1} \left(1 - \frac{i}{n}\right) \frac{c^k k^{k-2}}{k!} \left(1 - \frac{c}{n}\right)^{kn-k(k+3)/2+1} \\ &\sim \frac{1}{c} n \frac{c^k k^{k-2}}{k!} \exp\left(-\frac{k^2}{2n} - kc + \frac{ck^2}{2n}\right) = \frac{1}{c} n \frac{k^{k-2} (ce^{-c})^k}{k!} \exp\left(-\frac{(1-c)k^2}{2n}\right) \end{aligned} \quad (2.5)$$

For sufficiently k , one can apply Stirling's approximation to get an asymptotic estimate:

$$E_p^T(k) \sim \frac{n}{\sqrt{2\pi c}} k^{-5/2} (ce^{1-c})^k \quad (2.6)$$

Also, after multiplying each term by k , and summing from $k = 1$ to n and sending $n \mapsto \infty$, we obtain that the number of vertices in tree components are asymptotically:

$$E_p^T \sim n \cdot \frac{1}{c} \sum_{k=1}^{\infty} \frac{k^{k-1} (ce^{-c})^k}{k!} = t(c)n \quad (2.7)$$

Needless to say, $t(c)$ is decreasing in c for $c > 1$. Less trivially, $t(c) = 1$ for $c \leq 1$. This is because $s(c) = ct(c)$ is the only solution of $se^{-s} = ce^{-c}$, via a term-by-term series comparison. See [9] for details. Next, near criticality, i.e. $c = 1 + \epsilon$, $k < n^{2/3}$, and $|\epsilon| < 1/2$, one can easily show that the following upper bound holds:

$$E_p^T(k) \leq C \cdot nk^{-5/2} \exp(k(\log(1+\epsilon) - \epsilon) + \epsilon k^2/2n) \leq C \cdot nk^{-5/2} \exp\left(-\frac{\epsilon^2}{2}k + \frac{\epsilon^3}{3}k + \epsilon k^2/2n\right) \quad (2.8)$$

The above corollaries focused on bounding the number of (k, d) components in $G(n, p = (1+\epsilon)/n)$ graphs. One can obtain a far more general result by bounding the susceptibility of the random graph. Essentially, the i -th susceptibility of a random graph, up to a factor of n , is the i -th moment of the component size distribution of a vertex drawn randomly from the graph. Thus, upper bounds on higher order susceptibilities will give a bound on the variance of the number of components of a given size, which together with Chebyshev's inequality, will give concentration guarantees. The following theorem formalizes this intuition.

Theorem 3. *Let $X(k, d)$ be the number of components of type (k, d) . Define a subset $\Lambda \subset \mathbb{N} \times \{\mathbb{N} \cup \{0, -1\}\}$ to be the component types of interest, and let $-1 < \epsilon(n) \leq n/4$, $p = (1+\epsilon)/n$. Set $X_i = \sum \{k^i X(k, d) : (k, d) \in \Lambda\}$, $\mu_i = E[X_i]$, and $\tilde{k} = \max\{k : (k, d) \in \Lambda\}$. If $\epsilon \geq -(1+\epsilon)^2/n$ and $\tilde{k}^2(\epsilon + (1+\epsilon)^2/n) \leq n$, then $\text{Var}[X_i] \leq \mu_{2i} + \frac{2}{n} \left(\epsilon + \frac{(1+\epsilon)^2}{n}\right) \mu_{i+1}^2$, and if $-1 < \epsilon \leq -(1+\epsilon)^2/n$, then $\text{Var}[X_i] \leq \mu_{2i}$.*

Proof. We shall start with the following self-explanatory identities. Let $A = k_1 + k_2 - d_1 - d_2$, $B = (k_1 + k_2)(n - k_1 - k_2) + \binom{k_1+k_2}{2} - k_1 - k_2 - d_1 - d_2$.

$$\begin{aligned} E[X_i] &= \sum_{(k,d) \in \Lambda} k^i \binom{n}{k} C(k, d) p^{k+d} (1-p)^{k(n-k) + \binom{k}{2} - k - d} \\ E[X(k_1, d_1)X(k_2, d_2)] &= \binom{n}{k_1} \binom{n-k_1}{k_2} C(k_1, d_1) C(k_2, d_2) \cdot p^A (1-p)^B \\ &= E[X(k_1, d_1)] E[X(k_2, d_2)] \frac{\binom{n}{k_1+k_2}}{\binom{n}{k_1} \binom{n}{k_2}} (1-p)^{-k_1 k_2} \\ E[X(k, d)^2] &= E[X(k, d)] + E[X(k, d)]^2 \frac{\binom{n}{2k}}{\binom{n}{k}^2} (1-p)^{-k^2} \end{aligned} \quad (2.9)$$

where the last two identities are simply counting the number of expected pairs of (k, d) components, with the latter case including the case where the same component is selected twice. These identities imply:

$$\begin{aligned} E[X_i^2] &= E\left[\sum k^{2i} X(k, d)^2 + \sum k_1^i k_2^i X(k_1, d_1) X(k_2, d_2)\right] \\ &= E[X_{2i}] + \sum k_1^i k_2^i E[X(k_1, d_1)] E[X(k_2, d_2)] \frac{\binom{n}{k_1+k_2}}{\binom{n}{k_1} \binom{n}{k_2}} (1-p)^{-k_1 k_2} \end{aligned} \quad (2.10)$$

Next step are the inequalities: $\frac{(n)_{k_1+k_2}}{(n)_{k_1}(n)_{k_2}} = \prod_{i=0}^{k_1-1} \left(1 - \frac{k_2+i}{n}\right) / \left(1 - \frac{i}{n}\right) \leq \exp\left(\sum_{i=0}^{k_1-1} \frac{i}{n} - \frac{k_2+i}{n}\right) = \exp(-k_1 k_2/n)$.

For $(1-p)^{-k_1 k_2}$, as $p = (1+\epsilon)/n$, $1-p \geq \exp\left(-\frac{1+\epsilon}{n} - \frac{(1+\epsilon)^2}{n^2}\right)$. Putting these two together yield $\frac{(n)_{k_1+k_2}}{(n)_{k_1}(n)_{k_2}} (1-p)^{-k_1 k_2} \leq \exp\left(\frac{k_1 k_2}{n} \left(\epsilon + \frac{(1+\epsilon)^2}{n}\right)\right)$.

Hence, if $\epsilon + (1+\epsilon)^2/n < 0$, the above expression is less than 1, and hence $E[X_i^2] \leq E[X_{2i}] + E[X_i]^2$, and hence $\text{Var}[X_i] \leq E[X_{2i}] = \mu_{2i}$. If $\epsilon + (1+\epsilon)^2/n \geq 0$, and $\tilde{k}^2(\epsilon + (1+\epsilon)^2/n) \leq n$, then $\frac{k_1 k_2}{n}(\epsilon + \frac{(1+\epsilon)^2}{n}) \leq 1$, and as $1+2x \geq e^x$ in that range, we have:

$$\text{Var}[X_i] \leq \mu_{2i} + 2 \left(\frac{\epsilon}{n} + \frac{(1+\epsilon)^2}{n^2} \right) \sum k_1^{i+1} k_2^{i+1} E[X(k_1, d_1) X(k_2, d_2)] = \mu_{2i} + 2 \left(\frac{\epsilon}{n} + \frac{(1+\epsilon)^2}{n^2} \right) \mu_{i+1}^2 \quad (2.11)$$

□

2.3 Pre-transition: Sparse Forests

With the preliminary results behind us, let us begin by painting a basic picture of the random graph before the transition graph. Most of the vertices belong to tree components, of which the maximum size is at most $\log n$. Let us begin by proving the first assertion.

Theorem 4. *Let T be the number of vertices in a tree-component.*

1. *If $p = o(1/n)$, then $T = n$ almost surely.*
2. *If $p = c/n$ for $c \neq 1$, $E[T] \sim t(c)n = n \frac{1}{c} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (ce^{-c})^k$. In particular, $t(c) = 1$ for $c \leq 1$, and decreases in c .*

Proof. Both results proceed from a simple expectations calculation. If X_k is the number of k -cycles, then $E[X_k] = \binom{n}{k} \frac{(k-1)!}{2} p^k \leq (np)^k / 2k$, so if $np < c < 1$, $E[X_k] \leq c^k / 2k$, and hence $\sum_{k=3}^{\infty} E(X_k) \leq \sum_{k=3}^{\infty} c^k = O(1)$. The second identity follows from Corollary 2. □

Consequently, to obtain an upper-bound on the size of the largest component in that regime, it suffices to bound the size of the largest tree component. Bollobás gives a rather precise estimate based on Poisson approximations:

Theorem 5 (Subcritical Behavior). *If $p = c/n$, $c < 1$, $\alpha = c - 1 - \log c$, and set $k_0(s) = \frac{1}{\alpha} (\log n - \frac{5}{2} \log \log n + s)$. Then, denote $X(s)$ as the number of tree components of order at least $k_0(s)$. Then X is asymptotically Poisson with mean $\lambda_0 = \frac{1}{c\sqrt{2\pi}} \frac{\alpha^{5/2}}{1-e^{-\alpha}} e^{-s}$*

Corollary 3. *Setting s sufficiently large (in Bollobás' notation, $\omega(n) \mapsto \infty$) but weaker than $\log \log n$, one obtains the probability bound for the largest component size L :*

$$P(L \leq \frac{1}{\alpha} (\log n - 5/2 \log \log n + \omega(n))) \mapsto 0 \quad (2.12)$$

Proof. The proof of this result foreshadows the techniques used in later sections. First, we shall state without proof the following statement:

Lemma 1. *If X_i are non-negative integer valued random variables, convergence in moments imply convergence in distribution, as long as the moments are sufficiently small ($\lim E[X^r] r^m / r! \mapsto 0$)*

This is a discrete version of Carleman's theorem, which states that F is uniquely determined by its moments iff $\sum_{n=1}^{\infty} M_n^{-1/n} = \infty$ (i.e. its moments are small enough). For our purposes, we shall only need this result to prove convergence to a Poisson distribution:

Corollary 4. Denote $E_r(X)$ as the r -th factorial moment of X , i.e. $E[X(X-1)\dots(X-(r-1))]$. The Poisson r.v. has $E_r(X) = \lambda^r$, which satisfies the moment bound of the previous lemma. Hence, $\lim E_r(X_n) - \lambda^r = 0$ for all r implies $X_n \mapsto^D \text{Pois}(\lambda)$.

Thus, setting $Y = \sum_{k=k_0}^n T_k$, where T_k are the number of tree components of size k , suffices to show $E_r[Y] \mapsto \lambda_0^r$. By Corollary 2, $E[T_k] \sim \frac{n}{c\sqrt{2\pi}} k^{-5/2} (ce^{1-c})^k$. Hence, as $e^{-\alpha} = ce^{1-c} < 1$, for $k_2 = c_2 \log n$, the sum from k_2 to n is dominated by a geometric series, whose first term is $o(n^{-M})$, so we may drop the tail terms. Thus, suffices to show that for $Z = \sum_{k=k_0}^{k_2} T_k$, $E_r[Z] = \lambda_0^r$. The same bound applies as before:

$$E[Z] \sim \frac{n}{c\sqrt{2\pi}} \sum_{k_0}^{k_2} k^{-5/2} e^{-\alpha k} \sim \frac{n}{c\sqrt{2\pi}} \sum_{\ell=0}^{\infty} \left(\frac{\log n}{\alpha} \right)^{-5/2} e^{-\alpha(k_0+\ell)} = \lambda_0 \quad (2.13)$$

As for the higher factorial bounds, one can argue roughly that $E_r(T_k) \sim E[T_k]^r$ as follows: $E_r(T_k)$ is the expected number of ordered r -tuples of distinct tree components of size k . One can independently select these at random, and subtract the case in which there is an overlap between the r trees, of which the probability is minuscule. Then, one can extend to $Z = \sum T_k$, and argue $E_r[Z] \sim E[Z]^r$. $\square$

Hence, we have shown that for $p = c/n$, $c < 1$, the size of the largest component, which is a tree, is bounded asymptotically by $\frac{1}{\alpha} \log n$. (The convergence to a Poisson distribution in fact gives concentration around that value, so it is both an upper and a lower bound). To recap the proof, we relied on Cayley's theorem and moment estimation for the number of trees in the random graph for small values of p .

Remark 1. One can point out that the above result has been computed for $G(n, p)$, whereas our object of interest is $G(n, M)$, the Erdős-Rényi graph process. To apply the aforementioned results to the graph process, note that the property that most vertices belong to trees, with the size of the trees being small, is a monotone graph property (if G has the property and $H \subset G$, so does H .) Hence, one can apply the result for a slightly larger p , and then obtain analogous results for corresponding M . See Appendix B for details.

2.4 Phase Transition: the Giant Component

As t approaches $n/2$, from the random graph G_t , previously a sparse union of small trees, emerges a giant component of order ϵn . The key step in proving this result is the following theorem:

Theorem 6 (Component Gap, loosely stated). *There exists $s_0 = o(n)$, $t_0 = n/2 + s_0$. Then, as $n \mapsto \infty$, with probability 1, for $t_0 \leq t \leq n$, a component of G_t is either bigger than $n^{2/3}$, or smaller than $n^{2/3}/2$.*

Intuitively, shortly after $t = n/2$, there exists a gap in the sequence of component size. Thus, each component is either small (less than $n^{2/3}/2$), or large (greater than $n^{2/3}$). This is a precursor to the emergence of the dominant component among the “large components”.

From the component gap result, how does one prove the emergence of the giant component? Suppose we can show the persistence of the gap for t beyond n to all of time. This means that after time t_0 , all the components are small (i.e. smaller than $n^{2/3}/2$), or large (bigger than $n^{2/3}$). Furthermore, if two small components are joined via a random edge, its size is smaller than $n^{2/3}$. Hence, if this gap result holds for all time beyond t_0 , none of the small components can ever become large by combining with themselves: they continue to be incorporated by the large components, who themselves coalesce into one giant component shortly after t_0 .

To formalize the above intuition, the following is a loose summary of the proof of the emergence of the giant component.

1. We shall prove the component gap result for $n/2 + o(n) \leq t \leq n$, and then extend the component gap result beyond $t \leq n$. In fact, for $t \geq 3n/5$, there are no small components of order at least $100 \log n$.

2. We prove a precise estimate on the number of vertices in small components after $t = n/2$ (equivalently, the number of vertices belong to large components). All vertices belong to either unicyclic components ($o(n)$) or tree components ($\Theta(n)$).
3. As there are sufficiently many vertices belonging to large components, these components quickly coalesce into a giant component. [Erdős-Rényi Sprinkling Argument]

Hence, I shall organize the proof into the aforementioned three steps.

2.4.1 Step 1: Proving the Component Gap Theorem

Let us begin by restating the component gap theorem more precisely.

Theorem 7. *Let $s_0 = (2 \log n)^{1/2} n^{2/3}$ and set $k_0(s) = \lfloor (3 \log s - \log n - 2 \log \log n) n^2 / s^2 \rfloor$. Then if $t = n/2 + s$, $s_0 \leq |s| < n/2$, either $k < k_0(s)$ or $k > n^{2/3}$. In particular, as k_0 is decreasing in s and $k_0(s_0) = n^{2/3}/2$, for $t_0 = n/2 + s_0 \leq t \leq n$, the graph G_t has no component whose order is between $n^{2/3}/2$ and $n^{2/3}$.*

Proof. Let us assume $s > 0$, as the proof for the negative case proceeds similarly. Instead of showing directly the gap between $n^{2/3}/2$ and $n^{2/3}$, Bollobás employs a useful trick. By setting $k_1(s) = \min(\lfloor n^{2/3} \rfloor, 3k_0(s))$, note $k_0(s_0) < 1/2 n^{2/3}$, $k_0(s_0) \approx 1/2 n^{2/3}$, and hence $k_1(s_0) \approx \lfloor n^{2/3} \rfloor$, and then decrease monotonically in s .

I claim that rather than showing that no component exists of size $k_0(s)$ to $n^{2/3}$, it suffices to show that no component exists from $k_0(s)$ to $k_1(s)$. For s near s_0 , we have just reasoned that the two statements were equivalent, as $k_1(s_0) = n^{2/3}$. We can then proceed by induction. Suppose that there are no components of size $k_0(s)$ to $k_1(s)$ for the considered range of s , and also assume by the inductive hypothesis that no component exists from $k_0(s)$ to $n^{2/3}$ for $s \leq s'$.

Now, for the inductive step $s = s' + 1$, it is clear from the definition that $k_1(s + 1) < 2k_0(s)$. Hence, the $t + 1$ th edge cannot join two “small” components to create something larger than $k_1(s' + 1)$. Also, we have already assumed that there are no components of size $k_0(s' + 1)$ to $k_1(s' + 1)$. Finally, by inductive hypothesis, the remaining “large” components are already larger than $n^{2/3}$. Thus, it suffices to show the weaker statement that there are no components of size $k_0(s) \leq k \leq k_1(s)$.

It suffices to show that the expected number of components of order k_0 to k_1 from $t = t_0$ to n goes to 0.:

$$\sum_{t=t_0}^n \sum_{k=k_0(s)}^{k_1(s)} E_t(k) = \sum_{t=t_0}^n \sum_{k=k_0(s)}^{k_1(s)} \binom{n}{k} C(k, d) \binom{\binom{n-k}{2}}{t-k-d} / \binom{N}{t} = o(1) \quad (2.14)$$

However, for Erdős-Rényi graph processes $G(n, t)$, $\binom{\binom{n-k}{2}}{t-k-d} / \binom{N}{t}$ is rather cumbersome and difficult to bound. Bollobás’ original paper [7] does this via elementary inequalities, but the proof is laborious and involves casework. We shall instead follow [9]’s more streamlined proof, which uses an approximation by $G(n, p)$, $p = t / \binom{n}{2}$. In this case, the probability is simply $p^{k+d} (1-p)^{\binom{k-1}{2} - k - d + (n-k)k}$. However, the translation to $G(n, p)$ imposes a more stringent bound: one needs to show that the equivalent sum is $o(n^{-1/2})$ to conclude that the original sum is $o(1)$. See Appendix B for details.

By Corollary 1 and 2, suffices to show:

$$\sum_{t=t_0}^n \sum_{k=k_0(s)}^{k_1(s)} E_p(k) = o(n^{-1/2}) \implies \sum_{t_0}^n \sum_{k=k_0(s)}^{k_1(s)} n k^{-5/2} \exp\left(-\frac{\epsilon^2}{2} k + \frac{\epsilon^3}{3} k + \frac{\epsilon k^2}{2n}\right) = o(n^{-1/2}) \quad (2.15)$$

As $p = (1 + \epsilon)/n = 2t/n^2 = 1/n + 2s/n^2$, $k/n = o(s/n) = o(\epsilon)$, so the $\frac{\epsilon k^2}{2n}$ term is dominated by the $-\frac{\epsilon^2}{2} k$ term, and hence it suffices to bound:

$$\sum_{t=t_0}^n \sum_{k=k_0(s)}^{k=k_1(s)} n k^{-5/2} \exp(-0.499\epsilon^2 k + \epsilon^3 k/3) = n \sum_{t=n/2+s_0}^n \sum_{k_0(s)}^{k_1(s)} F(k) \quad (2.16)$$

For $\epsilon > 0.001 \iff t > 1.001n/2$, one can easily see $F(k+1)/F(k) < 1 - 10^{-7}$, so the sum over k is dominated (up to a constant factor) by the first term. Thus, setting $t_1 = 1.001n/2$, let us divide between $t > t_1$ and $t < t_1$: $n \sum_{t=t_1}^n F(k_0(s)) + n \sum_{t=t_0}^{t_1} \sum_{k=k_0(s)}^{k_1(s)} F(k)$.

For the first sum, if $t_1 \leq t \leq n$, $F(k_0(s)) \leq c' \cdot k_0(s)^{-5/2} \exp(-(3.9 - 8s/(3n)) \log n)$ via plugging in k_0 , and one can show the first sum is $o(n^{-3/2})$ for $s \leq n/2$, and hence multiplying back the n gives $o(n^{-1/2})$. For the second sum, for $t_0 \leq t \leq t_1$ and $k_0(s) \leq k \leq 3k_0(s)$ (by our definition of k_1), $F(k) \leq c'' k^{-5/2} \exp(-0.49\epsilon^2 k)$, and as $k = \Theta(k_0) \leq n^2/s^2 \log n$, we obtain:

$$\begin{aligned} n \sum_{t=t_0}^{t_1} \sum_{k=k_0(s)}^{k_1(s)} F(k) &\leq C_0 (\log n)^{-5/2} n^{-5} \sum_{t=t_0}^{t_1} s^5 \sum_{k=k_0(s)}^{k_1(s)} \exp(-0.49\epsilon^2 k) \\ &\leq o(1) \sum_{t=t_0}^{t_1} s^{-2} \sim n^{-2/3} = o(n^{-1/2}) \end{aligned} \quad (2.17)$$

where we continuously exploit the definition of k_1 and $k_0(s)$. $\square$

The component gap theorem allows us to classify components near $t = n/2$ as “small” (if $|C| \leq n^{2/3}/2$) or “large” (if $|C| \geq n^{2/3}$), as long as $t \leq n$. To extend the component size gap to $t \geq n$, one needs information regarding the largest “small” component. Let $L_S(G)$ (Largest Small) be the maximum size of a component smaller than $n^{2/3}$.

Theorem 8 (Largest Small Component). *Let $2n^{-1/3} < |\epsilon| < 1/4$, $p = (1 + \epsilon)/n$, and define:*

$$g_\epsilon(k) = \log n - \frac{5}{2} \log k + k(\log(1 + \epsilon) - \epsilon) + 2 \log(1/\epsilon) + k^2 \epsilon / 2n \quad (2.18)$$

Let $k_2(n) < n^{2/3}$ such that $g_\epsilon(k_2) \mapsto -\infty$. Then, $L_S(G) < k_2$ almost surely.

Proof. The first part, which is more immediately useful, follows quite quickly. By Corollary 1, suffices to show that the expected number of trees from k_2 to $n^{2/3}$ is 0, and by Corollary 2, suffices to show:

$$n \sum_{k=k_2}^{n^{2/3}} k^{-5/2} \exp(\epsilon k^2 / 2n + k(\log(1 + \epsilon) - \epsilon)) = o(1) \quad (2.19)$$

As the quadratic term $k^2/2n$ is dominated by the linear term as ϵ is fixed and $k < n^{2/3}$, and $\log(1 + \epsilon) - \epsilon < 0$, the above expression is dominated by a geometric series with ratio $(1 + \epsilon)^{-\epsilon} \approx 1 - \epsilon^2$ for small ϵ , say, $|\epsilon| < 1/4$. Hence, the sum is bounded above by the geometric series with ratio $1 - \epsilon^2$, which is ϵ^{-2} times the first term. Thus the number of trees of size k_2 to $n^{2/3}$ is bounded above by $c' n \epsilon^{-2} k_2^{-5/2} \exp(\epsilon k_2^2 / 2n + k_2(\log(1 + \epsilon) - \epsilon)) = c' \exp(g_\epsilon(k_2)) \mapsto 0$, as desired. $\square$

Corollary 5. *If $0 < \epsilon < 1/2$, $L_S(G) \leq 3 \log n \epsilon^{-2}$ almost surely.*

Proof. Suffices to show $g_\epsilon(3(\log n) \epsilon^{-2}) \mapsto -\infty$.

$$g_\epsilon(3(\log n) \epsilon^{-2}) \leq c + \log n - 3 \log 1/\epsilon - 5/2 \log \log n + 3 \log n \epsilon^{-2} (\log(1 + \epsilon) - \epsilon) + O(\epsilon^{-3} \log^2 n / n) \quad (2.20)$$

As $3\epsilon^{-2}(\log(1 + \epsilon) - \epsilon) \approx -3/2$, the above term is bounded above by $-1/2 \log n$, which goes to $-\infty$. $\square$

Corollary 6. *For $t \geq 3n/5$, G has no small component of order at least $100 \log n$.*

Proof. Here we deviate slightly from Bollobás' original proof written in [7] and [9]. In the original proof, Bollobás demonstrates that for a fixed $t \geq 5n/8$, the result holds, while we need a stronger moment bound to ensure that the result holds for all t . By the previous corollary, for $3n/5 < t < 5n/8$ (corresponding to $\epsilon = 1/5$ to $1/4$), the graph has no small component of order at least $75 \log n$. Now, for $a = 100 \log n$, as $2a < n^{2/3}$, if we prove that for $t \geq 5n/8$, there is no component of order between a and $2a$, we will have extended the component gap result to $t \geq n$.

First, we shall state a weak form of the connectivity-phase transition result for Erdős-Rényi graphs:

Lemma 2. *The Erdős-Rényi graph process is almost surely connected by $p = \frac{4 \log n}{n}$.*

We shall defer the simple proof to the appendix. Thus, suffices to show the gap between a and $2a$ for t from $5n/8$ to $2n \log n$. We shall show this by proving $\sum_{t=5n/8}^{2n \log n} E[\sum_{k=a}^{2a} \sum_{d=-1}^{\infty} X(k, d)] = o(1)$. Denote $2t/n = \alpha$. In much of the same argument as Corollary 1, one obtains:

$$\begin{aligned} E \left[\sum_{k=a}^{2a} \sum_{d=-1}^{k(k-1)/2-k} X(k, d) \right] &= \sum_{k=a}^{2a} \binom{n}{k} \sum_{d=-1}^{k(k-1)/2-k} C(k, d) \binom{\binom{n-k}{2}}{t-k-d} / \binom{N}{t} \\ &\leq \sum_{k=a}^{2a} c_2 n e^k k^{-5/2} \alpha^{k-1} e^{-2k(t-k)/n} \leq c_3 n \sum_{k=a}^{2a} e^{k-k\alpha} \alpha^k k^{-5/2} = c_3 n \sum_{k=a}^{2a} (\alpha e^{1-\alpha})^k k^{-5/2} \end{aligned} \quad (2.21)$$

Note that by assumption $\alpha > 5/4 = 1.25$, and the function $\alpha e^{1-\alpha}$ is decreasing from $\alpha > 1$. Hence, $\alpha e^{1-\alpha} < e^{1-1.25+\log(1.25)} < e^{-0.027}$. The above sum is thus bounded above by $c_3 n \cdot a \cdot a^{-5/2} e^{-0.027a} \leq c_4 n (\log n)^{-3/2} n^{-2.07} = o(n^{-1.07})$. Summing this bound over $t = 5n/8$ to $2n \log n$, we obtain that the expectation is $o(n^{-0.07} \log n) = o(1)$, and thus the component gap can be extended beyond $t = 5n/8$. $\square$

2.4.2 Step 2: Estimating the vertices in large components

Now that we have extended the component gap to all of $t \geq \frac{n}{2} + (2 \log n)^{1/2} n^{2/3}$, we need some precise bounds on the number of vertices in the small components shortly after $t = n/2$, and thereby estimate the number of vertices in the large components.

Theorem 9. *Let $(\log n)^{1/2} n^{-1/3} \leq \epsilon = o(1)$, and $p = (1 + \epsilon)/n$, and define $0 < \epsilon' < 1$ by $(1 - \epsilon')e^{\epsilon'} = (1 + \epsilon)e^{-\epsilon}$. Let Y_1 be the number of vertices on small (k, d) components with $d \geq 1$, Y_0 be the number of vertices on the small unicyclic components, and Y_{-1} be the number of vertices on the small tree components. Then, almost surely:*

1. $Y_1 = 0$
2. $Y_0 \leq \omega(n)\epsilon^{-2}$, for $\omega(n)$ any arbitrary function increasing to ∞ . (Loosely, $Y_0 \approx \epsilon^{-2}$)
3. $|Y_{-1} - \frac{1-\epsilon'}{1+\epsilon}n| \leq \omega(n)\epsilon^{-1/2}n^{1/2}$ As $\epsilon' = \epsilon - \frac{2}{3}\epsilon^2 + O(\epsilon^3)$, this implies:

$$|Y_{-1} - n + 2\epsilon n| \leq \omega(n)\epsilon^{-1/2}n^{1/2} + O(\epsilon^2 n) \quad (2.22)$$

Proof. First, by setting $k_\epsilon = 8 \log n \epsilon^{-2}$, we see that $g_\epsilon(k_\epsilon) \mapsto -\infty$, so there does not exist a component between the size k_ϵ and $n^{2/3}$. Hence, it suffices to consider $\tilde{Y}_1 = \sum_{k=4}^{k_\epsilon} \sum_{d \geq 1} kX(k, d)$, $\tilde{Y}_0 = \sum_{k=3}^{k_\epsilon} kX(k, 0)$, and $\tilde{Y}_{-1} = \sum_{k=1}^{k_\epsilon} kX(k, -1)$. As $k_\epsilon = O(n^{2/3})$, one can apply Corollary 1 and obtain:

$$E[Y_1] \leq c \sum_{k=4}^{k_\epsilon} k^{k+1} \binom{n}{k} \left(\frac{1+\epsilon}{n} \right)^{k+1} \left(1 - \frac{1+\epsilon}{n} \right)^{kn-k^2/2} \approx cn^{-1} \sum_{k=4}^{k_\epsilon} k^{1/2} \exp(-\epsilon^2 k/2) \mapsto 0 \quad (2.23)$$

For $E[Y_0]$, Theorem 39 shows that the number of unicyclic components are asymptotically $C(k, 0) \sim (\pi/8)^{1/2} k^{k-1/2}$, and hence:

$$\begin{aligned} E[Y_0] &\sim \sum_{k \leq k_\epsilon} k \binom{n}{k} (\sqrt{\pi/8})^{1/2} k^{k-1/2} \left(\frac{1+\epsilon}{n} \right)^k \left(1 - \frac{1+\epsilon}{n} \right)^{kn-k^2/2+3k/2} \\ &\sim \sum_{k \leq k_\epsilon} \exp(-k\epsilon^2/2) \sim \frac{1}{4} \int_0^\infty \exp(-x\epsilon^2/2) dx = \frac{1}{2} \epsilon^{-2} \end{aligned} \quad (2.24)$$

Thus, by Markov's inequality, $Y_0 \leq \omega(n)\epsilon^{-2}$ almost surely.

Finally, denote $E'[Y_i]$ as the expectation where $p = (1 - \epsilon')/n$. By the same calculation as above, one can show $E'[Y_1] = o(\epsilon^{-2})$, $E'[Y_0] \sim \frac{1}{2}\epsilon^{-2}$, and as $(1 - \epsilon')/n < 1$, our results from the subcritical section gives us that $E'[Y_{-1}] = n - \frac{1}{2}\epsilon^{-2} + o(\epsilon^{-2})$.

Now, to pass from $E'[Y_{-1}]$ to $E[Y_{-1}]$, note:

$$E[X(k, -1)]/E'[X(k, -1)] = \left(\frac{1+\epsilon}{1-\epsilon'} \right)^{k-1} \left(\frac{n-1-\epsilon}{n-1+\epsilon'} \right)^{kn-k^2/2-3k/2+1} = \frac{1-\epsilon'}{1+\epsilon} \exp(O(\epsilon k^2/n)) \quad (2.25)$$

Summing over $k \leq k_\epsilon$ $O(k^2/n)$ is insignificant, and $E[Y_{-1}] = \frac{1-\epsilon'}{1+\epsilon}n + O(\epsilon^{-2})$.

Secondly, to bound the variance of Y_{-1} , we have:

$$E \left(\sum_{k=1}^{k_\epsilon} k^2 X(k, -1) \right) = O(n) \sum_{k=1}^{k_\epsilon} k^{-1/2} \exp(-k\epsilon^2/2) \mapsto O(n) \int_1^\infty x^{-1/2} e^{-x\epsilon^2/2} dx = O(n\epsilon^{-1}) \quad (2.26)$$

Hence, by Theorem 3, $\text{Var}(Y_{-1}) = O(n\epsilon^{-1} + \epsilon/n(n\epsilon^{-1})^2) = O(n\epsilon^{-1})$. Consequently, by Chebyshev's inequality, with high probability,

$$|Y_{-1} - \frac{1-\epsilon'}{1+\epsilon}n| \leq \omega(n)n^{1/2}\epsilon^{-1/2} \quad (2.27)$$

□

As an upper bound on the number of vertices in small tree components is monotone, and similarly for other properties, one can translate the statement regarding p to a statement regarding t :

Corollary 7. *Suppose $t = (n/2) + s$ and $\frac{1}{2}(\log n)^{1/2}n^{2/3} \leq s = o(n)$, and $\omega(n) \mapsto \infty$. Then G_t has $4s + O(s^2/n) + O(\omega(n)n/s^{1/2})$ vertices on its large components, at most $\omega(n)n^2/s^2$ vertices on its small unicyclic components, the rest on its small tree components.*

Proof. We have shown that $Y_{-1} \approx (1 - 2\epsilon)n$, and hence $2\epsilon n$ belong in “large” components, if $p = (1 + \epsilon)/n$. In t space, this corresponds to $t = \frac{n^2}{2}p = \frac{(1+\epsilon)n}{2}$. Hence, the implied s is $\frac{\epsilon n}{2}$. Hence, $2\epsilon n = 4s$. The less important asymptotics follow similarly by translating the previous theorem into t and s space. □

2.4.3 Step 3: Erdős-Rényi Sprinkling Argument: Creation of the giant component

We are finally ready to prove the main result. Intuitively, once we have enough vertices belonging to large components, all of these components will coalesce quickly. This is intuitive: given two components of size N or greater, there exists at least N^2 edges between the components, and given sufficiently many vertices in these large components, this implies that all the large components will be connected almost surely only by adding a few edges. Such argument has been called the Erdős-Rényi sprinkling argument.

Theorem 10. *For every $t \geq t_1 = n/2 + 2(\log n)^{1/2}n^{2/3}$, the graph has a unique component of order at least $n^{2/3}$ and the other components have at most $n^{2/3}/2$ vertices each.*

Proof. As we have extended the component gap theorem beyond time $t \geq t_0 = n/2 + (2 \log n)^{1/2} n^{2/3}$, by the previous corollary, there exists at least $4\sqrt{2}(\log n)^{1/2} n^{2/3}$ vertices in the large components, and hence at t_0 one can partition the vertices belonging to the large components into subsets $V_1, \dots, V_m$, $m \geq 4\sqrt{2}(\log n)^{1/2}$, with $|V_i| \geq n^{2/3}$, with V_i belonging to one component.

Thus, suffices to show that V_1 through V_m become connected in a short amount of time. Setting $p = n^{-4/3}$, if one adds to G_{t_0} the remaining edges independently with probability p to obtain the graph H , almost surely one adds at most $t_0 + n^{2/3}/2 \leq t_1$. Thus, it suffices to show that V_1 through V_m are connected almost surely in H . The probability that V_i and V_j are connected is at least $1 - (1 - p)^{n^{4/3}} \geq 1 - e^{-1} \geq 1/2$.

Hence, the probability of V_i all being connected is bounded below by the probability that $G(m, 1/2)$ is connected, and as $m \mapsto \infty$, this is almost surely true, as desired. $\square$

Combining the above proof with Theorem 9, one obtains:

Corollary 8. *Let $t = n/2 + s$, with $2(\log n)^{1/2} n^{2/3} \leq s = o(n)$. Then whp the two largest component sizes satisfy $L_1(G_t) = 4s + O(s^2/n) + O(\omega(n)n/s^{1/2})$, and $L_2(G_t) \leq (\log n)n^2/s^2$.*

2.5 Post-Phase Transition: the Growth of the Giant Component

We have demonstrated the emergence of a giant component shortly after $t = n/2$. With the addition of each edge, the size of the giant component grows by roughly 4 vertices. Furthermore, the remaining components are quite small, with the second largest component being less than $(\log n)n^2/s^2$.

In fact, one can formulate an even stronger statement regarding the structure of the small components after $n/2$. We shall state the theorem without a proof, as the proof itself is very similar to the proof of the component gap theorem, relying on similar bounds and moment computations.

Theorem 11. *1. Small Components remain very small ($O(\log n)$): For $c_0 n/2 < t$, $1 < c_0$, define $t = cn/2$, and let $\alpha = (c - 1 - \log c)$. Then, almost surely G_t has no small component of order at least $\log n/\alpha$.*

2. Not that many unicyclic components: For $t > c_0 n/2$, every small component of G_t is either unicyclic or a tree, with at most $\omega(n) \log n$ unicyclic components.

3. Eventually no unicyclic components: for $\omega(n) \mapsto \infty$, if $t \geq \omega(n)n$, every component except the giant component is a tree.

In fact, by using the approximation $C(k, 0) \approx (\pi/8)^{1/2} k^{k-1/2}$, one can demonstrate that the expected number of unicyclic components of G_t having at most $2 \log n/\alpha$ is at most:

$$c \sum_{k=3}^{2 \log n/\alpha} \binom{n}{k} k^{k-1/2} \left(\frac{2t}{n^2}\right)^k \left(1 - \frac{2t}{n^2}\right)^{kn} \leq \sum_{k=3}^{2 \log n/\alpha} k^{-1} (ce^{1-c})^k = O(1) \quad (2.28)$$

and similarly for the number of vertices on the unicyclic components. Hence, there are at most $\omega(n)$ vertices on the unicyclic components.

Hence, to obtain a precise estimation of the order of the giant component, it suffices to obtain a precise estimate of the number of vertices in trees after the critical time $t = n/2$, i.e. $p = c/n$, $c > 1$, as the size of the giant component is simply the number of remaining vertices. By Corollary 2, $E[T] \sim n^{\frac{1}{c}} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (ce^{-c})^k = nt(c)$. Furthermore, by Theorem 3, $(\text{Var}[X_1] \leq \mu_1 + 3\epsilon\mu_2^2/n)$, one can bound the deviation from the expected norm and obtain the following theorem:

Theorem 12. *Given $p = c/n$, $c > 1$, $\omega(n) \mapsto \infty$, and $t(c) = \frac{1}{c} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (ce^{-c})^k$, then almost surely,*

$$|T(G) - t(c)n| < \omega(n)n^{1/2} \quad (2.29)$$

Hence, $L_1(G_t)$ is concentrated near $(1 - t(c))n$, within $O(n^{1/2})$ error. Furthermore, recall from the pre-phase transition section that the number of tree components of order $1/\alpha(\log n - 5/2 \log \log n + s)$ is asymptotically Poisson with mean $\lambda_0 = c(\alpha)e^{-s}$. Hence, one can analogously obtain concentration results for the i -th largest component for all i fixed. Summarizing, we have the following theorem:

Theorem 13 (Global Structure of the post-critical Erdős-Rényi process). *Let $c > 1$ be constant, $t = cn/2$, and $\omega(n) \mapsto \infty$. Then G consists of the giant component, small unicyclic components with $\omega(n)$ vertices, and small tree components. The order of the giant component satisfies:*

$$|L_1(G_t) - (1 - t(c))n| \leq \omega(n)n^{1/2} \quad (2.30)$$

where $t(c) = \frac{1}{c} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (ce^{-c})^k$, and for other fixed $i \geq 2$, given $\alpha = c - 1 - \log c$,

$$|L_i(G_t) - (1/\alpha) \left(\log n - \frac{5}{2} \log \log n \right)| \leq \omega(n) \quad (2.31)$$

Thus, with the exception of the giant component, the typical size of the small tree components remain relatively unchanged before and after the phase transition.

In contrast with the proliferation of small trees in the subcritical random graph, the giant component is far from a tree: as t becomes substantially larger than $n/2$, for instance, cn , with $c > 1/2$, as every component except the giant component is a tree, the complexity of the giant component (i.e. the number of edges beyond the number of vertices) grows at least linearly with respect to n .

Thus, the Erdős-Rényi random graph eventually simplifies to a linearly growing, increasingly complex, giant component, and a scattered forest of small tree components, which are inexorably absorbed by the giant component.

2.6 Conclusion and Renormalization

2.6.1 Brief Recap and Renormalization

To summarize Bollobás' seminal work, we have demonstrated that prior to the critical time $t = n/2$, the random graph process almost surely consists of isolated trees of order $O(\log n)$. Near the critical time $t = n/2 + O(n^{2/3} \log n^{1/2})$, there emerges a unique giant component of size $L_1(G)$, which then concentrates at $(1 - t(c))n$, with $t(c) = \frac{1}{c} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (ce^{-c})^k$. The remaining components, which are mostly trees, eventually become absorbed into the giant component until the graph is finally connected at $t = O(n \log n)$.

Anticipating future work regarding Achlioptas processes, one can renormalize the above results:

Theorem 14 (Precise evolution of the Giant Component). *There exists a function $\rho^{ER} : [0, \infty) \mapsto [0, 1)$ such that whenever $m/n \mapsto t$ as $n \mapsto \infty$, then $L_1(G(n, m))/n \mapsto^p \rho(t)$, where $\mapsto^p$ denotes convergence in probability. Then, for Erdős-Rényi processes, $\rho^{ER}(t) = 0$ for $t \leq 1/2$, $\rho(t) > 0$ for $t > 1/2$, and $\rho(t)$ is continuous at $t = 1/2$ with right-derivative 4 at this point. More precisely, for $t > 1/2$, $\rho(t) = 1 - \frac{1}{2t} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (2te^{-2t})^k$.*

Fig B.1 depicts the normalized ρ function and its derivative. As it can be read off the definition, $\rho(t)$ denotes the fraction of vertices belonging to the giant component.

While the results concerning the evolution of L_1 is of a global nature, one can also state the following “local” result concerning the number of components of order k . Denoting $N_k(G)$ as the number of vertices in components of degree k , recall from Theorem 13 that $N_k(G) = N_k^T(G) + o(n)$, where $N_k^T(G)$ is the number of vertices in trees of order k . Hence, by tree counting, and setting $p = (1 \pm \epsilon)/n$, we obtain:

$$N_k(G)/n \approx n^{-1} k^{k-1} \binom{n}{k} \left(\frac{1+\epsilon}{n} \right)^{k-1} \left(1 - \frac{1+\epsilon}{n} \right)^{nk-k^2/2} \approx \frac{1}{\sqrt{2\pi}} k^{-3/2} \exp\left(-\left(\frac{1}{2} + o(1)k\epsilon^2\right)\right) \quad (2.32)$$

Hence, renormalizing by setting $\rho_k(t) = N_k(G(nt))/n$, we obtain:

Theorem 15 (Small Components).

$$\rho_k(1/2 + \epsilon) \approx \frac{1}{\sqrt{2\pi}} k^{-3/2} e^{-(2+o(1))k\epsilon^2} \quad (2.33)$$

2.6.2 Weakly supercritical phase

One of the disadvantages of such a normalization is that we lose the detailed understanding of the microscopic dynamics around the critical percolation time $t = n/2$, also referred to as the weakly supercritical phase, e.g. in [25]. Note that in our proof, we did not actually prove that near the critical time the largest component is actually of size $\Theta(n^{2/3})$, although the component gap suggests $n^{2/3}$ is the correct magnitude.

Using a similar moment estimation technique, [38] demonstrates that for $t = n/2 + s$ with $sn^{-2/3} \mapsto \infty$, $L_1(G) \approx \frac{2(s+s')n}{n+2s}$, with $s' = O(s)$. Furthermore, the relative size of the components display an interesting dynamic. Initially, there are many components of relatively similar size, with the largest component changing regularly. Beyond a critical point, however, (roughly $t = n/2 + \Theta(n^{2/3})$), the largest component never switches, eventually becoming the giant component. Using a different technique involving multiplicative coalescents, Aldous [4] is able to show an even more precise result concerning the sequence of component orders rescaled by $n^{2/3}$.

These detailed results all concern the critical random graph, where we zoom into a neighborhood of the percolation time $t = n/2$, or equivalently $p = 1/n$. By studying the normalized functions ρ and ρ_k , we lose access to the granular results (those involving $n^{2/3}$ in Erdős-Rényi processes, for example). However, the loss of detail allows us to generalize the phase transition results to more general random graph processes.

In the next section, we shall describe the various generalizations of the Erdős-Rényi random graph process, such as preferential attachment processes and Achlioptas processes. While preferential attachment networks have been developed to provide models that incorporate some salient features of real-life networks, such as power-law degrees, Achlioptas processes are motivated by attempts to study the phase transition behavior.

After giving a brief overview of the various generalizations of the Erdős-Rényi graph processes, I shall focus on illuminating the phase transition dynamics of Achlioptas processes, as described in [26],[28], and [27]. In particular, I shall give an exposition of some of the techniques demonstrated in [25], which fully generalizes the phase transition behavior in Erdős Rényi graphs to general bounded-size Achlioptas processes.

Chapter 3

Generalization of Erdős-Rényi Random Graph Processes: a Brief Overview and Introduction to Achlioptas Processes

Before we move onto our main material on phase transition behavior of Achlioptas processes, let us slow down slightly to give a brief survey of various generalizations of the Erdős-Rényi random graph process, eventually working up to our main topic, Achlioptas processes.

The analysis of the canonical Erdős-Rényi random graph process has prompted several generalizations of the original graph model. While the exact forms and details of these models are myriad (see [19] for various applications), one can categorize many of the models based on their motivation: either the model seeks to better fit qualitative characteristics of real-world networks, or it seeks to test the robustness of Erdős-Rényi-like behavior in a more general setting.

3.1 Modelling Real World Networks

While the Erdős-Rényi model forms a reasonable null model for networks, real-life networks often exhibit qualitatively different behavior compared to the graph obtained by choosing m edges uniformly at random. An examination of real-world networks, such as social networks, connection of neurons, scientific co-authorship networks, and web-connections, reveals the following qualitative characteristics.

1. High-clustering behavior: intuitively, if v and w share a neighbor, they seem likelier to be connected. This is clearly in contrast with the Erdős Rényi model, where each edge is generated independently.
2. “A small world” network: the average distance between two randomly selected vertices is small. This is a characteristic shared by the Erdős-Rényi graph: for super-critical t (subcritical t does not make sense), the average distance between two vertices in the giant component is $O(\log n)$. Intuitively, this makes sense: a branching process modelling the exploration process from a random vertex v with mean $\lambda > 1$ grows roughly at size λ^t , and hence it reaches all n vertices in roughly $\log n / \log \lambda$ time. One can make this argument rigorous via an approximating branching process.
3. “Power-law” degree distribution: [3] claims that the degree distribution of real-world networks often follows a power-law distribution. In contrast, the marginal degree distribution in Erdős-Rényi models is roughly Poisson. with near independence across vertices. There has been some controversy over whether the degree distributions of real-world networks truly follow a power-law: see [11]

3.1.1 Clustering

To address these shortcomings of the Erdős-Rényi model, several alternative candidates have been introduced. To model clustering behavior, [23] introduces the latent-space model, where the vertices can be imagined to be embedded in a latent Euclidean space, with closer vertices likelier to be connected. Naturally, this model can better account for clustering behavior, although the dependence on an unknown latent space has raised identification issues.

One can also directly introduce clustering behavior by specifying the distribution of particular subgraphs in the network. These class of models, named exponential random graph models (ERGM), have received a lot of attention ([34]), mainly for their flexibility and their statistical interpretability. While the model parameters are readily interpretable as the prevalence of certain sub-structures in the network of interest, ERGM suffers from degeneracy [24]. As one can guess from the sharp phase transition thresholds, the frequency of various subgraphs tend to be sharply concentrated: for “most” parameter values, there are extremely few networks that satisfy the constraints, and are often radically different (the empty graph and the full graph both do not have any isolated triangles, for instance). Degeneracy and other issues typically make ERGM difficult to simulate.

Yet another model, known as inhomogeneous random graphs, was introduced by [10]. Instead of a uniform $p = c/n$, one considers a kernel Φ , a $n \times n$ matrix with $\Phi[i, j]$ the probability of an edge between vertex i and j . Taking n to the limit, one can consider a function $k : [0, 1]^2 \mapsto \mathbb{R}$, and let $p_{ij} = k(i/n, j/n)/n$. This will imply a vertex degree of $\Theta(1)$, which are often denoted as sparse graphs. By varying the function k , one can introduce clustering and power-law degrees to the random graph.

[18] introduces a similar idea, but with a different normalization. Here, the critical analogue of k is the graphon $W : [0, 1]^2 \mapsto [0, 1]$. Given a marginal distribution F over $[0, 1]$ and the graphon W , one can then simulate the random graph by sampling from F , and sampling each edge uniformly according to $W[i, j]$. Unlike the graph of [10], the expected degree is $\Theta(n)$, and hence the resulting graph is called “dense”. In the same way as inhomogeneous random graphs, one can introduce clustering through the notion of “latent communities”, i.e. have the graphon be a $k \times k$ diagonal-heavy block function, with each block representing a latent community, with members of the same community likelier to form a connection. The idea of decomposing a distribution into the marginals and pairwise relations resembles the idea of copula for scalar random variables. There is a wealth of statistical literature in estimating graphons: see, for example, [2].

3.1.2 Power-Law

To recreate a power-law degree distribution, [22] works directly with a random graph of a fixed degree distribution $p(d_i = k) = p_k$. To generate such a graph, one can generate d_i half-edges independently, and then pair these half-edges at random. While the resulting graph may have self-loops and multi-edges, there are a total of $O(1)$ of them, so they do not change the qualitative behavior of the graph.

These classes of model generate a rather simple phase-transition behavior regarding the emergence of a giant component. While it is easy to think that $\mu = \sum_k k p_k$ governs the emergence of the giant component (e.g. $\mu > 1$), note that given an exploration process, a vertex with degree k is k times as likely to be chosen than a vertex of degree 1. Hence, the distribution of the number of children of a first generation vertex and onward becomes $\tilde{p}_k \propto k p_k$. If we assume that p has finite second moment, then $\tilde{p}$ has finite mean $\tilde{\mu}$, and thus $\tilde{\mu}$ governs the emergence of a giant component.

However, the analogy between the Erdős-Rényi model does not extend fully: while $\tilde{\mu} > 1$ implies the emergence of the giant component, $\tilde{\mu} < 1$ does not imply that the largest component is of order $O(\log n)$. The precise asymptotics for $\tilde{\mu} < 1$ is in fact open: see [13].

While [22] directly works with a power-law process (or any general degree distribution), [3] derives a generative model for a graph process that yields a power-law degree distribution. The model, called the preferential-attachment model, introduces a new vertex and edges consecutively, with a connection between a high-degree vertex and the new vertex likelier than vertices of low-degree. A slight variation in the model parameter enables a power-law distribution with general power α .

3.2 How universal is the phase transition behavior of Erdős-Rényi Processes? Introduction to Achlioptas Processes

A second line of research regarding random-graph processes focuses on generalizing the qualitative behavior of Erdős-Rényi graphs to more general processes. In particular, one can ask whether the asymptotic behavior of L_1 , the size of the largest component, is qualitatively similar to Erdős-Rényi. For example, is there a t_c for which $L_1(G_t) = o(n)$ for $t < t_c$? Can there be a sudden jump in L_1 (an explosive percolation), or can one induce other significant qualitative changes in the phase transition? At a higher level, one can ask which of the features of the Erdős-Rényi phase transition are specific to the model, and which are “universal,” i.e. robust to perturbing the Erdős-Rényi model to a larger class of random graph processes.

In 2000, Dimitris Achlioptas introduced a candidate for such a generalization. His model, called Achlioptas processes, was generative, i.e. was defined as a random graph process rather than a single random graph. At each step, one randomly selects a fixed number of edges, and chooses to add to the graph a subset based purely on the component sizes of the selected edges. For example, if the rule is simply to add the first edge, we recover the Erdős-Rényi graph process. Another example, investigated by Bohman and Frieze [6], selects two edges, chooses the first edge if it connects two isolated vertices, and the second edge otherwise. As the process has a bias to connect isolated vertices, and hence delays the emergence of the giant component.

Achlioptas’ idea to perturb the random graph process at each time step, rather than varying the entire distribution G_t , has been motivated by the “power of two choices” paradigm introduced by [5] (see [20]). Imagine throwing n balls into n bins, with the goal of lowering the size of the maximal load. If one throws each ball uniformly at random, one can show that the maximum load is approximately $\log n / \log \log n$. [5] shows that if one at each step picks $d \geq 2$ bins uniformly at random, and then adds the ball to the bin with the minimal load, the resulting maximal load will be $\log \log n / \log d + \Theta(1)$ with high probability. Intuitively, the accumulation of simple choices over time can result in dramatic differences. Hopefully, Achlioptas reasoned, the same could hold for random graph processes.

3.2.1 Achlioptas Processes: A Formal Definition

Let us begin by giving a formal definition of Achlioptas processes:

Definition 1 (Achlioptas Processes). $\mathcal{G}_t$ is an ℓ -vertex Achlioptas process with rule R with the following generative process:

1. Select ℓ vertices $v_1, v_2, \dots, v_\ell$ uniformly at random.
2. Let $\vec{c} = (c(v_1), c(v_2), \dots, c(v_\ell))$ be the vector of component sizes of the selected vertices.
3. Let ℓ^{pair} be the set unordered pair of distinct elements in $[\ell]$, representing all edges between the selected ℓ vertices. Let the selection function $R : \vec{c} \mapsto \ell^{\text{pair}}$ be a deterministic rule selecting the edges to add only based on component sizes. Add the edges corresponding to $R(\vec{c})$.

Remark 2. The definition just given selects vertices uniformly at random rather than edges. Intuitively, these are almost equivalent. While the vertex selection rule may lead to a self-loop or multi-edges, one can show these never occur for tn steps with positive probability. Hence, if a certain phenomenon, such as local convergence, holds almost surely for the multi-graph, it continues to hold in the original graph case. Thus, it suffices to analyze the random vertex selection case, which is much more amenable to analysis.

By a similar logic, one can assume that one selects ℓ distinct vertices, and does not select edges already chosen (which happen with probability $1 - O(1/n)$.)

Among the possible edge selection rules, the product rule, which selects the edge that minimizes the product of the component sizes, received particular attention for drastically altering the phase transition behavior. While Achlioptas, D’Souza and Spencer [1] conjectured based on numerical evidence that the normalized $L_1(G_{tn})/n$ is discontinuous at t_c (a phenomenon named “explosive percolation”), Riordan and Warnke [26] disproved the conjecture, showing that all Achlioptas phase transitions were continuous. Their surprising result demonstrates that much has yet to be learned about the properties of Achlioptas processes.

3.2.2 Bounded-Achlioptas Processes: The “Universality” of Erdős-Rényi Phase Transition

While the behavior of general Achlioptas processes has resisted rigorous analysis, more is known regarding the behavior of bounded-size Achlioptas processes.

Definition 2. An Achlioptas process is “bounded-size” if there is $K < \infty$ such that the selection rule R treats components of size K or greater equally. Rigorously, let $\vec{c}_K = (\min\{\vec{c}_i, K\})$. Then, for all component sizes $\vec{c}$, $R(\vec{c}) = R(\vec{c}_K)$. The product-rule obviously is not a bounded-size rule.

In [25], Riordan and Warnke have established that “bounded-size” Achlioptas processes exhibit “Erdős-Rényi” like phase-transition behavior, and hence is a proper extension of the Erdős-Rényi process with respect to its phase-transition behavior. To be more specific, they proved results regarding L_1 , the size of the largest component, and N_k , the number of vertices in components of size k , and showed that they are qualitatively similar to the Erdős-Rényi process.

Size of the largest component

First, for $t < t_c$, i.e. the subcritical period, all components are uniformly small. Recall Theorem 5 in the Erdős-Rényi section, which states that L_r is concentrated at $\frac{1}{\alpha}(\log n - \frac{5}{2} \log \log n + O(1))$, with $\alpha = c - 1 - \log c$, $c = np$. Analogously, for bounded-size Achlioptas processes,

Theorem 16 ($t < t_c$: Subcritical Period). *Let R be a bounded-size ℓ -vertex rule with critical time $t_c > 0$. For $i/n = t_c - \epsilon$, $\epsilon \in [0, \epsilon_0]$ with $\epsilon^3 n \mapsto \infty$, we have the following upper bound for L_r , the size of the r -th largest component:*

$$L_r(i) = \psi(t_c - \epsilon)^{-1} \left(\log(\epsilon^3 n) - \frac{5}{2} \log \log(\epsilon^3 n) + O(1) \right) \quad (3.1)$$

Next, recall Corollary 8 which states that during the phase transition, i.e. $t = n/2 + s$, $2(\log n)^{1/2} n^{2/3} \leq s = o(n)$ i.e. the largest and the second largest component of the Erdős Rényi process satisfy:

$$L_1(G_t) = 4s + O(s^2/n) + O(\omega(n)n/s^{1/2}) \implies \rho(t_c + \epsilon) = 4\epsilon + O(\epsilon^2) \text{ and } L_2(G_t) \leq (\log n)n^2/s^2 \quad (3.2)$$

The bounded-size Achlioptas process satisfies the same property: the largest component (normalized) grows linearly post t_c , with the second largest component dominated by the giant component.

Theorem 17 ($t \geq t_c$: Phase Transition and Supercriticality). *Let R be a bounded-size ℓ -vertex rule. Then, $L_1(tn)/n \mapsto \rho(t)$ in probability, where ρ is 0 for $t < t_c$, $\rho(t_c) = 0$, and $\rho(t)$ analytic on $[t_c, t_c + \delta]$, with the right derivative of ρ at t_c strictly positive. In other words, for $\epsilon \in [0, \epsilon_0]$, $\rho(t_c + \epsilon) = \sum_{j=1}^{\infty} a_j \epsilon^j$, $a_1 > 0$.*

Furthermore, if $\epsilon = i/n - t_c$ satisfies $\epsilon^3 n \geq \omega$, $\omega \mapsto \infty$, $\tau = (\log \omega)^{-1/2}$ we have concentration around ρ , simultaneously for all i :

$$L_1(i) = (1 \pm \tau)\rho(t_c + \epsilon)n, \quad L_2(i) \leq \tau L_1(i) \quad (3.3)$$

To reemphasize, the above concentration holds simultaneously for all i outside the critical window $t \approx t_c n$, which is a much stronger statement than convergence for a fixed i !

Distribution of small components

Along with $L_1(G)$, we are also interested in $N_k(G)$, the number of vertices in components of size k , and its concentration to ρ_k . Recall Theorem 15, which shows that for Erdős-Rényi processes, $N_k/n \mapsto \rho_k$, and:

$$\rho_k(1/2 + \epsilon) \approx \frac{1}{\sqrt{2\pi}} k^{-3/2} e^{-k \cdot (2 + o(1))\epsilon^2} \quad (3.4)$$

The following theorem shows that the exact same result holds for bounded-size Achlioptas processes, with $\theta(t)$ instead of $\frac{1}{\sqrt{2\pi}}$ and $\psi(t)$ instead of $2\epsilon^2$:

Theorem 18. *For bounded-size ℓ -vertex rule Achlioptas processes, assuming all possible component sizes k can be reached, we have for $t \in [t_c - \epsilon_0, t_c + \epsilon_0]$:*

$$\rho_k(t) = (1 + O(1/k))k^{-3/2}\theta(t)e^{-\psi(t)k} \quad (3.5)$$

Moreover, convergence of N_k to ρ_k is “uniform” in k and t near $t = t_c$: for all $1 \leq k \leq n^{1/10}$ and $\epsilon = i/n - t_c$ satisfies $10\psi''(t_c)\epsilon^2k \leq \log n$ ($t \approx t_c$) k , one simultaneously has:

$$\begin{aligned} \frac{N_k(i)}{n} &= (1 \pm n^{-1/30}/k) \cdot \rho_k(i/n) \\ \frac{N_{\geq k}(i)}{n} &= (1 \pm n^{-1/30}/k) \cdot \left(\sum_{j \geq k} \rho_j(i/n) + \rho(i/n) \right) \end{aligned} \quad (3.6)$$

In particular, if we zoom in on $t \approx t_c$, at criticality $N_{\geq k} = (1 + O(1/k))Bk^{-1/2}n$, which implies a polynomial decay of the tail of the component size distribution.

For the remainder of the expository paper, we shall concentrate on understanding the qualitative results, as well as the key techniques, of Riordan and Warnke’s phase transition results for Achlioptas processes.

3.2.3 Key Techniques

In contrast to the classical results of [9] on Erdős-Rényi processes, which relied on clever applications of basic combinatorial bounds, the powerful results of [25] rely on several techniques developed over the past 20 years. The main difficulty lies in the fact that the edges are no longer independent: the addition of certain edges in the past has an impact in the distribution of edges added in the future. Hence, merely counting the number of possible graphs have limited applicability in these more general random graph processes.

To analyze Achlioptas processes, one therefore needs a better understanding of two key techniques: the branching-process approximation and the differential equation method.

We have already encountered an example of the branching process technique, which involves pairing the original random graph process with an approximate branching process. Recall that in our heuristic argument for why $p = 1/n$ is the critical time for the Erdős-Rényi process, we approximated the neighborhood exploration process (where we pick a previously unexplored vertex v , reveal all of its neighbors, add them to the stack, and remove v from the stack) with a Poisson branching process.

The benefit from introducing the auxiliary branching process is that compared to the original graph process, the branching process is typically more tractable via the use of generating functions (see [15] for examples). The difficulty, of course, arises from bounding the estimation errors from the original random graph process and the branching process. [13] has many examples of such analyses.

The second technique, called the Differential Equation Method (DEM), uses differential equations to compute the evolution of certain quantities of interest in random graph processes. Wormald’s key theorem, which appeared in [36], shows that under certain conditions, a measurable quantity Y_t of the random graph process G_t become sharply concentrated around its expected value, which follows a differential equation.

Intuitively, DEM is far more powerful for generalized random graph processes than the standard ER processes. This is because in contrast with Erdős-Rényi, we often have very limited understanding of the distribution of the random graph G at a fixed time t . As is the case for preferential attachment or general Achlioptas processes, we are only given the generative process of the random graph. Thus, while the properties of the random graph at any fixed time t may be intractable, one can hope to ascertain the evolution of $E[Y_{t+1}|Y_t]$ from the generative process. By studying the implicit differential equations, either analytically or numerically, one can often gain a better understanding of the original graph process.

These two techniques have been developed to analyze not only Achlioptas processes, but also preferential attachment models, Newman-Strogatz graphs, etc. Consequently, not only are they imperative to understanding the universality result for bounded-size Achlioptas processes, they also merit independent importance as a crucial technique for analyzing random graph processes in general. Hence, we shall devote the next sections to give a basic exposition of the two techniques, before we finally move onto the universality result of [25].

Chapter 4

Technique 1: Branching Processes

The first main technique we shall explore involves modelling the neighborhood exploration process in random graph processes via a random graph process. We have already seen a heuristic sketch done for the Erdős-Rényi model, where the offspring distribution is $\text{Bin}(n, p)$.

For our purposes, we not only need the parameter p_c where there is percolation, but also more precise results regarding the exact number of particles (for information regarding component size distributions), as well as survival probability (for estimating the giant component). We shall first develop the basic theory, in particular establishing bounds on the total number of particles for the subcritical case and the survival probability for the supercritical case. The necessary condition for these bounds involve subexponential tails for the offspring distribution, or almost equivalently, a uniform bound on $g_X(\beta)$ for $\beta > 1$, where g_X is the probability generating function for the offspring distribution.

For our purposes, we need to consider more complicated versions of the classical branching process. To exploit the fact that the Achlioptas processes are bounded, [25] considers the neighborhood exploration process, treating vertices in large (greater than K) components differently from vertices in small components. Furthermore, the vertices in small components are treated as known, with only the vertices in large components “revealed” as part of the exploration process. Thus, the relevant branching process naturally has two types of particles: the vertices in the large components, which yield offsprings, and the vertices in the small components, which do not. Hence, in order to analyze the approximating branching process, one needs to extend the theory of branching processes to this special type, which [16] names “sesqui-type processes”.

While the new process is certainly more difficult to analyze, the spirit of the result is the same. First, the survival probability of the sesqui-type process depends only on the distribution of the fertile offsprings. Hence, no new analysis is needed for the survival probability of supercritical sesqui-type process. As for bounds on the total progeny of a subcritical sesqui-type process, a similar result still holds: if the distributions of both types of offspring are “jointly” subexponential, i.e. the probability generating function $g(\alpha, \beta)$ is well-defined for $|\alpha|, |\beta| < R$, $R > 1$, then the probability mass function for total progeny decays exponentially. In fact, we will be able to show far more accurate asymptotics near criticality.

The proof, which appears in [16], is complicated, and involves careful estimation of a contour integral. We shall give a sketch of the proof in Appendix C. For our purposes, we shall cite the main theorem, and give an intuition to why this is a generalization of the original branching process result.

Finally, we shall conclude by introducing the notion of a t_c -critical branching process family, which is a family of branching processes whose parameters vary analytically in t , with the branching process reaching criticality at $t = t_c$.

4.1 Single-Variable Branching Processes: Criticality at $\mu = 1$

I shall first give a quick review of the single-variable branching process. For convenience, denote $\mathbb{N} = \{0, 1, 2, \dots\}$, the set of non-negative integers. Let F be a probability mass function over $\mathbb{N}$, which is the distribution of the number of children.

4.1.1 Basic Definition and Results

Definition 3 (Galton-Watson Discrete Time Branching Process). *Let X_n be the number of active particles at time n . Given X_n , $X_{n+1} = Y_1 + Y_2 + \dots + Y_{X_n}$, where Y_i are i.i.d. F . The process continues until $X_n = 0$, i.e. extinction occurs. Denote $T_n = X_1 + \dots + X_n$ to be the total number of particles up to time n .*

The main tool to analyze branching processes is the probability generating function (PGF) $g_X(s) = \sum_{n=0}^{\infty} P(X = n)s^n$. Denote g_F as the PGF of F , and g_X be the PGF of X_n , the number of active particles.

$$g_{X_{n+1}} = \sum_n P(X_n = n) E[s^{Y_1 + Y_2 + \dots + Y_n}] = \sum_n P(X_n = n) g_F^n(s) = g_X(g_F(s)) \quad (4.1)$$

If we begin with one particle, i.e. $X_1 = 1$, we thus have the PGF for the n -th generation particles is the n -th iteration of g_F : $g_{X_n} = g_F^{(n)}(s)$. By Dominated Convergence Theorem, one can exchange summation with differentiation, and thus the PGF gives us the moments of a distribution, provided they are finite:

$$\begin{aligned} \frac{dg}{ds} &= \sum_{n=0}^{\infty} n P(X = n) s^{n-1} \implies g'(1) = E[X] \\ \frac{d^2g}{ds^2} &= \sum_{n=0}^{\infty} n(n-1) P(X = n) s^{n-2} \implies \text{Var}[X] = g''(1) - (g'(1))^2 - g'(1) \end{aligned} \quad (4.2)$$

Applying this to the PGF for the n th generation, it suffices to compute the first and second derivative of $g^{(n)}(s)$. By chain-rule, one obtains the following recursive equation:

$$\begin{aligned} \frac{d}{ds} g^{(n)}(s) &= g'(g^{(n-1)}(s)) (g^{(n-1)}(s))' \\ \frac{d^2}{ds^2} g^{(n)}(s) &= g''(g^{(n-1)}(s)) \left((g^{(n-1)}(s))' \right)^2 + g'(g^{(n-1)}(s)) (g^{(n-1)}(s))'' \end{aligned} \quad (4.3)$$

Plugging in $s = 1$, one obtains the mean and variance of F :

$$\begin{aligned} \mu_n &= \mu \cdot \mu_{n-1} \implies \mu_n = \mu^n \\ \sigma_n^2 &= \mu^2 \sigma_{n-1}^2 + \mu^{n-1} \sigma^2 \implies \sigma_n^2 = \mu^{n-1} \sigma^2 \frac{1 - \mu^n}{1 - \mu} \text{ (or if } \mu = 1 \text{), } n \sigma^2 \end{aligned} \quad (4.4)$$

Finally, of particular interest is the survival probability: $P(T_{\infty} = \infty)$. As for $\mu < 1$, by linearity of expectation, one can obtain $E[T_{\infty}] = \frac{1}{1-\mu} < \infty$, and hence one obtains almost sure extinction. What about $\mu \geq 1$? Note that $P(X_n = 0) = g^{(n)}(0) = g(g^{(n-1)}(0)) = g(P(X_{n-1} = 0))$. Furthermore, trivially $P(X_n = 0) \geq P(X_{n-1} = 0)$. Hence, the probability of extinction by time n , denoted as θ_n , is a monotonically increasing sequence satisfying $\theta_n = g(\theta_{n-1})$. Thus, there exists a limit, which satisfies $\theta = g(\theta)$. By construction of g , $g(1) = 1$, g is convex and increasing, with $g'(1) = \mu$. Hence, we obtain the following:

1. (Subcritical Case): For $\mu \leq 1$, the only solution is at $\theta = 1$, and hence extinction is certain.
2. (Supercritical Case): For $\mu > 1$, there exists a unique smaller solution $\theta < 1$, the desired probability.

4.1.2 Subcriticality: Bounds on Total Population

Oftentimes, one can argue that the offspring distribution F is subexponential, i.e. there exists c and K such that $P(X \geq s) \leq K \exp(-cs)$. Then, one can obtain subexponential bounds for the total population T for the subcritical case $\mu < 1$.

Theorem 19. *Let X be the offspring distribution with $E[X] = \mu < 1$, and suppose there exists K and c such that $P(X \geq s) \leq K e^{-cs}$. Then, the following hold:*

1. The distribution of T , the total population, is also subexponential.
2. There exists $\beta > 1$ such that for $1 \leq x \leq \beta$, there exists a uniform bound on the probability generating function $F_T(x) = E[x^T] \leq D$.

Proof. First, say $\lambda > 0$ is smaller than c , e.g. $\lambda = c/2$. Then, one can bound $\exp(\lambda X)$. Being rather loose with the notation, I will assume X can be reasonably represented by a continuous density f on $[0, \infty)$ -an argument which can be made precise by taking sequences f_n that converges in distribution to f_X :

$$E[\exp(\lambda X)] = \int \exp(\lambda x) f(x) dx = -(1 - F(x)) \exp(\lambda x) \Big|_0^\infty + \lambda \int (1 - F(x)) \exp(\lambda x) dx \leq \frac{\lambda K}{c - \lambda} \quad (4.5)$$

Next, we introduce the random walk associated to the branching process: $S_t = S_0 + \sum_{i=1}^t (X_i - 1)$, with the stopping time T being the hitting time of 0 of the random walk S_t , which corresponds precisely to the total size of the branching process. As $\mu < 1$, we know $T < \infty$ with probability 1.

From the above bound on $E[\exp(\lambda X)]$, one can demonstrate that the following is a supermartingale:

$$M_t = \exp(\lambda S_t) \cdot \left(\frac{(c - \lambda)e^\lambda}{\lambda K} \right)^t \quad (4.6)$$

Hence, by applying optional stopping, one obtains:

$$E[M_0] = \exp(\lambda) \geq E[M_T] = E \left[\left(\frac{(c - \lambda)e^\lambda}{\lambda K} \right)^T \right] \quad (4.7)$$

One can let λ sufficiently small such that $\frac{(c - \lambda)e^\lambda}{\lambda K} = \beta > 1$, and one obtains $E[\beta^T] \leq \exp(\lambda) = D$, which is the desired uniform bound on the generating function. To conclude, one can simply apply Markov/Chernoff, i.e. $P(T \geq s) \leq E[\beta^T] \beta^{-s} \leq D \beta^{-s}$, to obtain subexponentiality of T . $\square$

4.1.3 Bounds on Survival Probability

For the supercritical case $\mu > 1$, one can obtain bounds on the survival probability, i.e. $P(T_\infty = \infty)$. The key technical idea comes from taking second-order Taylor approximations of the generating function. We shall first prove an easy lower bound on the survival probability, before moving onto a stronger result.

Theorem 20. *Let the offspring distribution X have $E[X] = \mu > 1$ and $\text{Var}[X] = \sigma^2$.*

1. *Given s independent copies of X , $P(\sum_{i=1}^s (X_i - 1) \leq -1) \leq \beta^s$ for some $\beta < 1$. This implies the probability that the process will go extinct at time $T > s$ decays exponentially at rate β . (Hence, extinction, if it ever happens, is the likeliest early on.)*
2. *The Galton-Watson process terminates with probability at most $y < 1$, where β, y depends only on μ, σ .*

Proof. Let us look at the pgf of $X - 1$, i.e. $\phi(t) = E[e^{-t(X-1)}]$. Easily, $\phi(0) = 1$, $\phi'(0) = \mu - 1$, and $\phi''(t) = E[(X - 1)^2 e^{-t(X-1)}] \leq P(X = 0)e^t + E[X^2] \leq e + \mu^2 + \sigma^2$ for $0 \leq t \leq 1$, so $\phi(t) \leq 1 + (\mu - 1)t + (e + \mu^2 + \sigma^2) \frac{t^2}{2}$. As $\mu > 1$, one can either minimize the quadratic or set $t = 1$ to obtain $\phi(t) < 1$.

The first statement then follows from a standard Chernoff argument:

$$P\left(\sum_{i=1}^s X_i \leq s - 1\right) \leq P\left(\sum_{i=1}^s (X_i - 1) \leq -1\right) \leq \phi(t)^s \quad (4.8)$$

As for the probability of extinction, note from the above analysis that the probability of extinction θ is the minimal nonnegative solution to $\theta = F_X(\theta)$. As $F_X(1) = 1$, $F'_X(1) = \mu$ and $F''_X(1) < \sigma^2 + \mu^2$, hence $F_X(1 - t) < 1 - t\mu + (\sigma^2 + \mu^2) \frac{t^2}{2}$, and in particular $t = 2(\mu - 1)/(\sigma^2 + \mu^2)$, $F_X(1 - t) < 1 - t$. Hence $\theta < 1 - 2(\mu - 1)/(\sigma^2 + \mu^2)$, as desired. $\square$

The basic argument made above can be extended to obtain more precise results about the survival probabilities near criticality ($E[X] = 1$). Recall from the single-type section that the extinction probability θ is the smallest non-negative solution to $\theta = g_X(\theta)$, by letting $\rho = 1 - \theta$ be the survival probability, we obtain:

$$g_X(1 - x) = E(1 - x)^X = \sum_{n=0}^{\infty} (-1)^n E\left[\binom{X}{n} x^n\right] \quad (4.9)$$

By setting $h_X(x) = \frac{1 - g_X(1 - x)}{x}$, we obtain an analytic function on the domain of g_X , with:

$$\begin{aligned} h(x) &= EX - E\binom{X}{2}x + E\binom{X}{3}x^3 + \dots \\ h(0) &= E[X] = g'(1), \quad -h'(0) = E\binom{X}{2} = \frac{E(X(X-1))}{2} \end{aligned} \quad (4.10)$$

From the definition, ρ satisfies $h_X(\rho) = 1$, and hence intuitively one can use an inverse function theorem of h_X near 1 to obtain results concerning ρ . Unsurprisingly, we shall need some bounds on the derivative of h_X to apply the inverse function theorem. The following lemma, from [16], makes the argument rigorous:

Lemma 3 (Survival Probability Near Criticality). *Suppose $R > 1$, $g_X(R) \leq M$, and there exists δ and k_1 such that $\delta \leq \min\{P(X = k_1), P(X = k_1 + 1)\}$. Then, the following hold.*

1. *There exists a c_x such that for $|x| \leq c_x$, for each $m \in \mathbb{N}$ there exists $c^{(m)}$ such that $|D^m h_X(x)| \leq c^{(m)}$.*
2. *If $E[X] \geq 1 - \delta$, $P(X \geq 2) > 0$, and $h'_X(0) \leq -\delta$, and if $|x| < c_x$, $h'_X(0) \leq -\delta/2$.*
3. *There is a constant ϵ_x such that if $|EX - 1| < \epsilon_x$, there is a unique $\hat{\rho} = \hat{\rho}_X \in \{|x| < c_x\}$ such that $h_X(\hat{\rho}) = 1$.*
4. *Furthermore, $\rho_X = \max\{\hat{\rho}, 0\}$, $\text{sign}(\hat{\rho}) = \text{sign}(EX - 1)$, and $|\hat{\rho}| = \Theta(|EX - 1|)$.*

Proof. As $g_X(R) \leq M$ and $h_X = (1 - g_X(1 - x))/x$, if $|x| < (R - 1)/2$, $h_X = O(1)$, and one can use Cauchy estimates to bound the derivatives for $|x|$ sufficiently small. Next, one can show that $E[X(X - 1)] \geq 2\delta$, and as $h'_X(0) = -\frac{1}{2}E[X(X - 1)]$, the bound follows. The result for $|x| < c_x$ follows from the bound on the second derivative.

Next, let $F(x) = h_X(x) - EX$, where $F(0) = 0$. As $F'(0) \leq -\delta/2$, one can find near 0 an inverse function $G(y)$, by the inverse function theorem. Next, define $\hat{\rho} = G(1 - EX)$, which is the desired inverse value for 1. As $\|DG\| = O(1)$, and $\|DF\| = O(1)$ in their respective domains via the bound above, one can easily establish $|\hat{\rho}| = \Theta(|EX - 1|)$.

To conclude the proof, one needs to relate $\hat{\rho}$ with the survival probability ρ_X . Recall from the basic fixed point analysis of g_X , a convex function, that there are at most two solutions in the interval $[0, R]$. By definition, 1 and $1 - \hat{\rho}$ are solutions, and as G has domain on a sufficiently small neighborhood of 1, one can show $1 - \hat{\rho} \in (0, R)$. If $\hat{\rho} > 0$, then we immediately have the two distinct solutions $1 - \hat{\rho}$ and 1, and hence $\hat{\rho} = \rho_X$. If $\hat{\rho} \leq 0$, then $1 - \hat{\rho} \in (1, R)$, and hence $1 - \rho_X = 1$, and $g'_X(1) < 1$ by strict convexity, so $EX - 1 < 0$, and so forth. $\square$

To summarize the lemma, we have shown that under certain regularity conditions, near $EX = 1$, the survival probability is 0 for $EX < 1$, and increases roughly linearly $\hat{\rho} = \Theta(|EX - 1|)$ as $EX \geq 1$. As a fruit of our effort, we have not only provided a lower bound for survival probability, but have demonstrated locally linear growth.

To anticipate the development needed later, one can use this to obtain bounds on the survival probability of a branching process family near criticality. Before we introduce the notion of a branching process family near criticality, we need to introduce a slight generalization of the original Galton-Watson branching process: now, each particle can also give birth to a separate type of offspring which are barren.

4.2 Sesqui-type Processes and Branching Process Families

To understand the estimation in [25], we need a slight variant of the canonical branching process described above, called the sesqui-type branching processes, formulated in [16]. As can be gleaned from the prefix “sesqui”, a sesqui-type branching process has one and a half type of particles. What does it mean to have a half-type? It means that there are two types of particles, L and S , where only L have offsprings and S are barren.

4.2.1 Basic Definition and Results

Hence, to describe a sesqui-type branching process, it suffices to specify the starting distribution for the two types of particles, (Y^0, Z^0) , as well as the distribution of (Y, Z) , which is the number of offspring of each type.

Definition 4. *Given a probability distribution (Y, Z) on $\mathbb{N}^2$ (the “offspring” distribution of types (L, S)), and the initial distribution (Y^0, Z^0) , the branching process S_{Y,Z,Y^0,Z^0} is defined as follows: given Y^0 particles of type L and Z^0 particles of type S , each L particle gives birth independently to Y L particles and Z S particles. S particles have no children. We shall write $|\mathcal{S}|$ to be the total number of particles.*

For ease of analysis, we shall assume a regularity condition on the distribution of Y, Z , specifically a uniform bound on $g_{Y,Z}(R)$ for $R > 1$ (roughly equivalent to the subexponential bound in previous sections.)

Definition 5. *Let $R > 1$, $M < \infty$, $k_1, k_2 \in \mathbb{N}$, $\delta > 0$.*

1. $\mathcal{K}^0(R, M, \delta)$ *is the set of probability distributions on $\mathbb{N}^2$ such that $E[R^{Y+Z}] \leq M$, $E[Y] \geq \delta$.*
2. $\mathcal{K}^1(k_1, k_2, \delta)$ *is the set of distributions $\pi_{i,j}$ on $\mathbb{N}$ such that $\pi_{k_1, k_2}, \pi_{k_1, k_2+1}, \pi_{k_1+1, k_2} \geq \delta$.*
3. $\mathcal{K} = \mathcal{K}^0 \cap \mathcal{K}^1$.

Intuitively, $\mathcal{K}^0$ imposes a uniform bound on the probability generating function, and $\mathcal{K}^1$ ensures that (Y, Z) is not essentially supported on a sublattice. Also, $(Y, Z) \in \mathcal{K}^1$ trivially implies $EZ \geq \delta$ and $EY \geq \delta$.

Then, the key result is the following theorem:

Theorem 21 (Asymptotic Point Probability of $\mathcal{S}$). *For $(Y^0, Z^0) \in \mathcal{K}^0$, $(Y, Z) \in \mathcal{K}$, and $|E[Y] - 1| \leq c_1$ with $c_1 > 0$ a constant, then for all $N \geq 1$, we have:*

$$P(|\mathcal{S}| = N) = N^{-3/2} e^{-N\xi} (\theta + O(N^{-1})) \quad (4.11)$$

where $\xi = -\Psi(x^*) \geq 0$, $\theta = \sqrt{2\pi/|\Psi''(x^*)|} \Phi(x^*) = \Theta(1)$, and $\xi = \Theta(|EY - 1|^2)$.

While the formula seems formidable, the key takeaway is that for the near critical case of $EY = 1 \pm \epsilon$, $P(|\mathcal{S}| = N)$ decays exponentially in $\Theta(\epsilon^2 N)$. Note that if one fixes R, M, k_1, k_2 , and δ , the above approximation is uniform in N ! Thus, this is an analogue of Theorem 19, which states that given subexponential bounds on offspring distribution X , the total progeny T is also subexponential. We shall give a sketch of the proof in Appendix C.

4.2.2 Branching Process Families

For our purposes, we wish to generalize the asymptotic formula slightly to allow the distribution of Y, Z to differ over time, indexed by t . Let us define the notion of a t_c -critical branching process family.

Definition 6. *Let $t_0 < t_c < t_1$. The branching process family $\mathcal{S}_t$, $t \in (t_0, t_1)$ is t_c -critical if the following hold:*

1. **Analyticity:** *There exist $\delta > 0$ and $R > 1$ such that $g_t(y, z)$ and $g_t^0(y, z)$ are defined and analytic on the domain $(t_c - \delta, t_c + \delta) \subset (t_0, t_1)$ and $|y|, |z| < R$.*

2. t_c **critical, and process increasing in t** : $E[Y_{t_c}] = 1$, $E[Y_{t_c}^0] > 0$, and $\frac{d}{dt}E[Y_t]|_{t=t_c} > 0$

3. **Spanning/non-periodicity**: There exists $k_0 \in \mathbb{N}$ such that:

$$\min \{P(Y_{t_c} = k_0, Z_{t_c} = k_0), P(Y_{t_c} = k_0, Z_{t_c} = k_0 + 1), P(Y_{t_c} = k_0 + 1, Z_{t_c} = k_0)\} > 0$$

The analyticity condition is a simple regularity condition, and the third condition guarantees that (Y, Z) does not concentrate on a sublattice of $\mathbb{N}^2$. More importantly, the second condition suggests that if $t < t_c \implies E[Y_t] < 1$, we are at a subcritical phase, while $t > t_c$ implies supercriticality, with the phase transition occurring at $t = t_c$. For our purposes, the machinery of t_c -critical branching processes allows us to model a random graph process near the phase transition.

Next, one can show that the critical quantities $P(|\mathcal{S}| = N)$, the point probability, and ρ , the survival probability, vary smoothly in t . By walking through the proof of Theorem 40, [16] shows that if $g_u(y, z)$ and $g_u^0(y, z)$ are analytic in u, y , and z jointly, then the critical quantities θ_u and ξ_u vary smoothly. Specifically, we obtain:

Lemma 4 (Continuity for Branching Process Family: Point Probability). *Suppose $R > 1$, $M < \infty$, $k_1, k_2 \in \mathbb{N}$, $\delta > 0$. Set $\mathcal{K}^0, \mathcal{K}$ as before, and let $(Y_u, Z_u) \in \mathcal{K}$ and $(Y_u^0, Z_u^0) \in \mathcal{K}^0$ for all $u \in I$. Critically, let $|E[Y_u] - 1| \leq c_1$ for all $u \in I$.*

Let $\theta_u = \Theta(1)$ and $\xi_u = \Theta(|EY - 1|^2)$ be the parameters in Theorem 40 such that:

$$P(|\mathcal{S}_u| = N) = N^{-3/2} e^{-N\xi_u} (\theta_u + O(N^{-1})) \quad (4.12)$$

If $g_u(y, z)$ and $g_u^0(y, z)$ are analytic as functions of $(u, y, z) \in I \times \{(y, z) \in \mathbb{C} : |y|, |z| < R\}$, and $\max\{|\frac{\partial}{\partial u} g_u|, |\frac{\partial}{\partial u} g_u^0|\} \leq \lambda$ for some λ , then ξ_u and θ_u are analytic as functions of $u \in I$. Furthermore, $\frac{d}{du} \xi_u = O(\lambda |EY_u - 1|)$, $\frac{d}{du} \theta_u = O(\lambda)$.

Simply put, given g_u locally constant (a uniform bound on its derivative w.r.t. u), the probability bound of $N^{-3/2} e^{-N\xi} \theta$ continues to hold with ξ and θ being roughly constant (the derivatives bounded).

In fact, not only do the point probability bounds stay relatively constant, but so do the survival probabilities: suppose that we are near criticality sufficiently such that $|E[Y_u] - 1| < \epsilon_x$ in lemma 3. Let $\hat{\rho}$ (loosely the inverse image of 1 of h_X) be defined as before.

Lemma 5. $\hat{\rho}_u$ is also analytic, with $\frac{d}{du} \hat{\rho}_u = O(\lambda)$

We shall omit the proof of this result, as it follows from the definition of $\hat{\rho}$ and basic Cauchy estimates on the derivatives.

Chapter 5

Technique 2: DEM

The idea behind the differential equation method (DEM) is simple. Oftentimes, we do not have a clear understanding of the distribution and the related quantities of a random process Y_t , measurable with respect to a filtration F_t . However, we often know how Y_{t+1} evolves conditioned on Y_t . Thus, one can often formulate a difference equation involving the expectation of Y_t , e.g. equations of the form:

$$E[Y_{t+1} - Y_t | F_t] = f(t, Y_t) + o(1) \quad (5.1)$$

Then, one naturally dreams of turning the difference equation into a differential equation: under the normalization $z(x) = Y_{nx}/n$, $z(x)$ should satisfy $z'(x) = f(x, z(x))$. Thus, by examining the differential equation, either analytically or numerically, one obtains the value for the expectation of Y_t . Of course, the convergence of Y_t to z depends on critical regularity assumptions, such as f being Lipschitz, the deviation from f being $o(1)$, and the process Y_t satisfying bounded differences. These conditions are not merely technical—failure of these conditions often indicate real limits in the stochastic process, such as percolation.

We shall begin our discussion of DEM by introducing some (not very rigorous) toy examples to motivate Wormald's theorem.

5.1 Toy Examples

5.1.1 Balls and Bins

We shall begin with a basic combinatorial process. Imagine there being n empty bins, and we continuously throw balls uniformly at random into the bins. Let $X(m)$ be the number of empty bins after throwing m balls. This is naturally a Markov process, and one can easily compute the expectation of $X(m)$ to be $n \cdot (1 - \frac{1}{n})^m$. For $t = m/n$, the expectation converges to $n \cdot e^{-t}$. Furthermore, Hoeffding-Azuma gives us the concentration of the martingale $Y_m = E[X(T) | F_m]$, where $|Y_m - Y_{m-1}| \leq 1$, so $P(|X(T) - E[X(T)]| > \sqrt{\alpha T}) < 2e^{-\alpha/2}$.

Speaking loosely, one can derive this result in another way. Note that the expectation of $X(i+1) - X(i)$ given $X(i)$ is -1 with probability $\frac{X(i)}{n}$ and 0 otherwise. Hence,

$$E[X(i+1) - X(i) | X(i)] = -\frac{X(i)}{n} \quad (5.2)$$

Under the normalization $x(t) = \frac{1}{n}X(tn)$, the above equation becomes:

$$x(t + \frac{1}{n}) - x(t) = -\frac{1}{n}x(t) \implies dx/dt = -x \quad (5.3)$$

Thus, the solution to the differential equation is $x(t) = Ce^{-t}$, and initial conditions yield $x(t) = e^{-t}$, which matches our original result.

Perhaps we are interested in not just the number of empty bins, but also the number of bins with exactly one ball, which we shall denote as $Y(i)$. Then, we are looking at a system of equations:

$$\begin{aligned} E[X(i+1) - X(i)|F_i] &= -\frac{X(i)}{n} \rightarrow dx/dt = -x \\ E[Y(i+1) - Y(i)|F_i] &= -\frac{Y(i)}{n} + \frac{X(i)}{n} \rightarrow dy/dt = -y + x \end{aligned} \quad (5.4)$$

The differential equations yield the solution $x(t) = e^{-t}$, $y(t) = te^{-t}$. One can easily verify that the expectations indeed converge to these functions via elementary calculations. The argument generalizes to the joint evolution of $X_0, X_1, \dots, X_n$ for n finite.

5.1.2 Independence number of $G_{n,r}$

The second example from [12] involves estimating the independence number of $G_{n,r}$, an r -regular random graph, i.e. an r -regular graph drawn uniformly at random across all possible r -regular graph. Recall that the independence number of a graph is the maximal number of vertices which do not neighbor each other.

Finding the maximum independent set in a graph, even for r -regular graphs, is NP-hard. One can, however, hope to find a lower-bound via the following simple greedy algorithm:

1. Initialize $S = \emptyset$, $T = V$, the vertex set.
2. While $T \neq \emptyset$, choose $v \in T$ uniformly at random, and add it to S . Then, reveal all the r edges from v to other vertices, and then remove v and all of its neighbors from T .
3. Output $|S|$.

One can visualize the greedy algorithm as an evolution of the following random graph process. We start with the empty graph with n vertices. At each step, we add to the graph the edges that are revealed by the algorithm. We keep track of Y_m , the number of vertices with zero degree, which are still available for the greedy algorithm to add to S .

Naturally, $Y_0 = n$, and $|Y_{m+1} - Y_m| \leq r + 1$ (so Hoeffding-Azuma still applies). As for the evolution of the expectation of Y_t , we first have at each step the vertex that is selected into S . What about the remaining vertices? Note that for r -regular graphs, one can imagine nr ordered pairings representing each edge twice. At time m , we have $nr - 2mr$ remaining ordered pairs. Among them, rY_m pairs contain a yet unselected vertex. Hence, the probability that one of the r revealed edges is connected to a vertex of degree 0 is $\frac{Y_m}{n-2m}$, and by linearity of expectation, we have $\frac{rY_m}{n-2m}$ expected additional lost vertices. Summarizing, we obtain:

$$E[Y_{m+1} - Y_m|G_m] = -1 - \frac{rY_m}{n-2m} \quad (5.5)$$

This yields the following differential equation, for $z(t) = Y(nt)/n$:

$$\frac{dz(x)}{dx} = -1 - \frac{rz(x)}{1-2x} \quad (5.6)$$

Setting $y = 1-2x$, we obtain the lovely o.d.e. $rz(y) - 2yz'(y) + y = 0$, which by standard methods (integrating factors: [33]) yield:

$$z(x) = C(1-2x)^{r/2} - \frac{1}{r-2}(1-2x) \quad (5.7)$$

Setting $z(0) = 1$ gives $C = \frac{r-1}{r-2}$. Solving for $z(x) = 0$ gives $\frac{1}{2} \left(1 - \left(\frac{1}{r-1} \right)^{2/(r-2)} \right)$, which is the greedy estimation for the independence number of an r -regular random graph.

Of course, the argument above, for both the balls and bins and the greedy algorithm analysis, is only suggestive: we did not formally prove that the random process Y_t concentrates around a deterministic function ρ , which then follows a differential equation governed via the expectations equation: $E[Y_{t+1} - Y_t|F_t]$. The following theorem, by Wormald [36], shows that this intuitive fact in fact is true.

5.2 Wormald's Theorem

Before we state and prove Wormald's theorem, I shall first state some basic inequalities regarding random variables/processes with finite differences. They are immediately useful in the proof of Wormald's theorem, and also are an essential tool in the analysis of combinatorial processes, where bounded differences arise in a very natural way. We shall defer the proofs of these famous inequalities to Appendix D.

Theorem 22 (Hoeffding-Azuma's Inequality). *Suppose X_n is a martingale with respect to a filtration F_n satisfying the bounded differences condition with $P(|X_n - X_{n-1}| \leq c_n) = 1$. Then, for any $n > 0$, we have:*

$$P(|X_n - X_0| > t) \leq 2 \exp \left(-\frac{t^2}{2 \sum_{i=1}^n c_i^2} \right) \quad (5.8)$$

Corollary 9 (McDiamard's Inequality). *Suppose $f : \mathbb{R}^n \mapsto \mathbb{R}$ is Lipschitz with respect to the Hamming metric, i.e. if x and x' only differ in the k -th coordinate, there exists $\sigma_k > 0$ such that $|f(x) - f(x')| \leq \sigma_k$. Then, given $Y = f(X_1, \dots, X_n)$ for X_i independent and any $\alpha > 0$:*

$$P(|Y - EY| \geq \alpha) \leq 2 \exp \left(-\frac{2\alpha^2}{\sum \sigma_i^2} \right) \quad (5.9)$$

Let us now state Wormald's theorem, following the exposition given in [37].

Theorem 23. *Suppose the random variables Y_i , $1 \leq i \leq K$, are measurable with respect to a discrete-time Markovian process G_t . Assume $D \subset \mathbb{R}^{K+1}$ is bounded, connected, open, and contains the set $\{(0, z_1, \dots, z_K) : \Pr[Y_i(0) = z_i, 1 \leq i \leq K] \neq 0 \text{ for some } n\}$. Intuitively, D has to contain all the points corresponding to initial conditions. Let $t < T_D$. We need three critical conditions to be satisfied:*

1. (Boundedness): For some $\beta(n) \geq 1$ and $\gamma(n)$, and $t < T_D$

$$P(\max |Y_i(t+1) - Y_i(t)| \leq \beta |G_t| \geq 1 - \gamma) \quad (5.10)$$

2. (Trend): For $\lambda_1(n) = o(1)$, for all i we have:

$$|E(Y_i(t+1) - Y_i(t) | G_t) - f_i(t/n, Y_1(t)/n, \dots, Y_K(t)/n| \leq \lambda_1 \quad (5.11)$$

3. (Lipschitz) Each f_i is continuous and Lipschitz on $D \cap \{(t, z_1, \dots, z_K) : t \geq 0\}$.

Then, the following hold:

1. For $z' \in D$, $\frac{dz_i}{dx} = f_i(x, z_1, \dots, z_K)$ has a unique solution in D satisfying the initial conditions $z(0) = z'$ extending arbitrarily close to the boundary of D .
2. For $\lambda > \lambda_1 + C_0 n \gamma(n)$, $\lambda = o(1)$. For C sufficiently large, with probability $1 - O(n\gamma + \beta/\lambda \exp(-\frac{n\lambda^3}{\beta^3}))$,

$$Y_i(t) = n z_i(t/n) + O(\lambda n) \quad (5.12)$$

uniformly for $0 \leq t \leq \sigma n$, where σ is the supremum of x such that $z(x)$ can be extended before reaching the distance $C\lambda$ of the boundary of D .

Proof. First, by a standard theorem in the theory of ODE (the Cauchy-Lipschitz theorem), as f are Lipschitz with the same constant, the differential equations have a unique solution.

We shall prove the rest of the theorem for $K = 1$, as the proof of the general case is identical. The key idea is to use Hoeffding-Azuma to bound the deviation $Y(t+k) - Y(t) - k f(t/n, Y(t)/n)$. Let $\omega = n\lambda/\beta$, and suffices to assume $\beta/\lambda > n^{1/3}$ or else the probability bound is vacuous. Thus, $\omega > n^{2/3}$. Similarly, let $\lambda < 1$. We wish to show the concentration of $Y(t+\omega) - Y(t)$. Note:

$$E[Y(t+k+1) - Y(t+k) | G_{t+k}] = f((t+k)/n, Y(t+k)/n) + O(\lambda_1) = f(t/n, Y(t)/n) + O(\lambda_1 + k\beta/n) \quad (5.13)$$

where we have used the boundedness hypothesis $|Y(t+k) - Y(t)| \leq k\beta$, and f being Lipschitz. Thus, there exists a function $g(n) = O(\lambda_1 + \omega\beta/n) = O(\lambda)$ such that $Y(t+k) - Y(t) - kf(t/n, Y(t)/n) - kg(n)$ is a supermartingale in k .

Furthermore, the differences in this supermartingale are at most $\beta + O(1) \leq c\beta$. Hoeffding-Azuma thus gives us a bound on the upper tail, and a similar argument using a submartingale can bound the lower tail. Thus, for all $\alpha > 0$:

$$P(|Y(t+\omega) - Y(t) - \omega f(t/n, Y(t)/n)| \geq \omega g(n) + c\beta\sqrt{2\omega\alpha}|G_T) \leq 2e^{-\alpha} \quad (5.14)$$

The rest of the proof is routine, where one seeks to bound, for $k_i = i\omega$, $i \leq \sigma n/\omega$, and $j \leq i$

$$P(|Y(k_j) - z(k_j/n)n| \geq B_j \text{ for some } j \leq i) = O(ie^{-\alpha}) \quad (5.15)$$

We can prove this by induction. Naturally, $Y(k_{i+1}) - z(k_{i+1}/n)n$ can be decomposed into $Y(k_i) - z(k_i/n)n$, $Y(k_{i+1}) - Y(k_i) - \omega f(k_i/n, Y(k_i)/n)$, $\omega z'(k_i/n) + z(k_i/n)n - z(k_{i+1}/n)n$, and $\omega f(k_i/n, Y(k_i)/n) - \omega z'(k_i/n)$. We can bound these components separately, the first via the inductive hypothesis, the rest via 5.14 and Lipschitz conditions. As $B_i = O(n\lambda)$, setting $\alpha = n\lambda^3/\beta^3$ gives us the theorem. $\square$

At a high level, the idea of transforming the stochastic process of interest into a martingale and applying Hoeffding-Azuma seems widely applicable, even where Wormald's theorem does not strictly apply.

5.3 Applications to Achlioptas Processes: Small Components and Susceptibility

5.3.1 Local Convergence

To conclude our introduction to DEM, we shall apply the techniques to prove local convergence of bounded-size Achlioptas processes. For clarity, we shall focus our attention on the 4-vertex Achlioptas processes, although the argument readily extends to ℓ vertex rules. Recall that $N_k(G)$ is the number of vertices in components of size k in G . We define two notions of the convergence of the Achlioptas process G_t :

Definition 7 (Convergence of G_t). 1. A graph process G_t is locally convergent if $N_k(G_{nt})/n \mapsto^p \rho_k(t)$.
2. G_t is globally convergent if $L_1(G(tn))/n \mapsto^p \rho(t)$, where $L_1(G)$ is the size of the largest component.

Employing the differential method, Wormald and Spencer [29] demonstrate that all bounded-size Achlioptas processes are locally convergent. Loosely following their exposition, we shall give a quick proof of their results regarding local convergence. The significance of local convergence comes from [26], which shows that local convergence implies global convergence of G_t . Thus, the convergence of N_k up to normalization gives us information about the emergence of the giant component in Achlioptas processes.

The rule that defines the Achlioptas process is dependent on $(c(v_1), c(v_2), c(v_3), c(v_4))$, i.e. the component size of each of the 4 vertices. Let $\Omega = \{1, 2, \dots, K, \omega\}$, where ω stands for all components bigger than size K . The Achlioptas process is bounded-size because it treats all components of size greater than K the same.

The Achlioptas rule can be formulated as follows: given $F \subset \Omega^4$, the graph adds the first edge $\{v_1, v_2\}$ if $c(\vec{v}) \in F$, and $\{v_3, v_4\}$ otherwise. The step is called redundant if the resulting added edge has both vertices in the same component. For example, the Bohman-Frieze process is defined by $F = \{(1, 1, \alpha, \beta), \alpha, \beta \in \Omega\}$. Simply, if the first edge joins two isolated vertices, add it to the graph, and if not, add the second edge. The Erdős-Rényi process simply sets $F = \Omega^4$.

Let $R_k(t) = \frac{N_k(t)}{n}$. Then, as we are drawing vertices uniformly at random, $P(c(\vec{v}) = \vec{j}) = R_{j_1}R_{j_2}R_{j_3}R_{j_4}$. Denote $\Delta(\vec{j}, k)$ to be half of the change in N_k when $c(\vec{v}) = \vec{j}$. Say, WLOG, $\vec{j} \in F$, and hence we join $\{v_1, v_2\}$. (The other case where the second edge is selected can be dealt symmetrically) There are three cases concerning $\Delta(\vec{j}, k)$:

1. , $j_1, j_2 \neq \omega$: the rule joins two “small” components of size j_1 and j_2 . If $j_1 \neq j_2$, $\Delta(\vec{j}, j_i) = -\frac{1}{2}j_i$, or $-j$ if $j_1 = j_2 = j$. If $j_1 + j_2 > K$, $\Delta(\vec{j}, \omega) = \frac{1}{2}(j_1 + j_2)$, otherwise $\Delta(\vec{j}, j_1 + j_2) = \frac{1}{2}(j_1 + j_2)$.
2. $j_1 = \omega$, $j_2 \neq \omega$ (and symmetrically $j_2 = \omega$, $j_1 \neq \omega$). Then, a large component absorbs a small component, $\Delta(\vec{j}, \omega) = \frac{1}{2}j_1$, $\Delta(\vec{j}, j_1) = -\frac{1}{2}j_1$.
3. $j_1, j_2 \in \omega$: then no changes are made.

The point is that $|\Delta| \leq K$, which will yield the boundedness condition for Wormald’s theorem.

Denote $R_i^*(t+1) = R_i + \frac{2}{n}\Delta_i$, and $e_i = R_i(t+1) - R_i^*(t+1)$. Intuitively, if the step is not redundant (i.e. joins two vertices of different components), $e_i = 0$, and otherwise is $\frac{2}{n}\Delta_i$. (Hence, it captures the error term from assuming that all steps are non-redundant). Note that for $e_i \neq 0$, Δ_i should not be 0, which implies that the component containing the selected edge must have size at most K , which occurs with probability at most $2Kn^{-1}$, and $|e_i| \leq \frac{2}{n}|\Delta| \leq \frac{2K}{n}$. Thus, $E[e_i]/(2/n) \leq 2K^2n^{-1}$. Thus, the following describes the trend condition:

$$\frac{E[R_k(t+1) - R_k(t)|G_t]}{2/n} = \sum_{\vec{j}} \Delta(\vec{j}; k) R_{j_1} R_{j_2} R_{j_3} R_{j_4} + O(K^2n^{-1}) \quad (5.16)$$

Also, $|\Delta| \leq K$, we obtain the boundedness condition:

$$|R_k(t+1) - R_k(t)| \leq 2Kn^{-1} \quad (5.17)$$

The difference equation above translates to the following system of differential equations:

$$\frac{d}{dt}\rho_k = \sum_{\vec{j} \in \Omega^4} \Delta(\vec{j}; k) \rho_{j_1} \rho_{j_2} \rho_{j_3} \rho_{j_4} \rho_1(0) = 1; \rho_i(0) = 0 \quad (5.18)$$

The differential equation 5.18 shall appear quite frequently, so it is good to have a basic analysis.

Theorem 24. $\rho_i(t)$ is defined for all $t \geq 0$, with $\sum_i \rho_i(t) = 1$, $\rho_i > 0$ and $\rho'_\omega(t) > 0$ for all $t > 0$.

Proof. As we basically have an ordinary differential equation $\vec{\rho}' = f(\vec{\rho})$ with $f \in C^\infty$, by Picard’s theorem, there is a solution in some neighborhood. Note that for all $\vec{j}$, $\sum_{i \in \Omega} \Delta(\vec{j}, i) = 0$, so summing over the differential equations over i gives us $\sum_i \rho_i(t)$ is constant at 1.

Furthermore, one can prove that the first non-zero derivative of x_i at 0 is positive (this includes the zeroth-derivative, which is the function value). The proof follows by basic induction and the observation that ρ'_i contains the term $i\rho_{i-1}\rho_1\rho_{i-1}\rho_1$ as this necessarily introduces a new component of size i .

Hence, as the first nonzero derivative is positive, $\rho_i(t) > 0$ in $(0, \epsilon)$. Now, suppose for the sake of contradiction that there exists a minimal t such that $\rho_i(t) = 0$. Fixing $0 < t' < t$, for $[t', t)$, as all of the terms are positive and add to 1, the negative terms appearing in ρ'_i sum to at most $-C\rho_i$, and hence ρ_i is bounded below by an exponential function, which contradicts $\rho_i(t) = 0$, as desired.

The rest follows quickly. The polynomial for $\rho'_\omega(t)$ only has positive coefficients, and as $\rho_i > 0$, the sum is always positive, as desired. As ρ_i are always nonnegative and sum to 1, $\rho_i \in [0, 1]$ whenever defined, and hence $\vec{\rho}' = f(\vec{\rho})$ takes place in a compact space, so $\vec{\rho}(t)$ is defined for all $t \geq 0$, as desired. $\square$

With the analysis of the differential equation finished, we are ready to prove local convergence:

Theorem 25. Let $\vec{R}(t) = (R_1(G_t), R_2(G_t), \dots, R_K(G_t), R_\omega(G_t))$. Let $\rho_k(t)$ be the solution to the differential equation 5.18. Then, $P(|\vec{R}(i) - \vec{\rho}(2i/n)|_\infty = O(n^{-1/4}) = 1 - O(\exp(-n^{1/5}))$, uniformly for $0 \leq i \leq \tau n/2$ for any fixed τ .

Proof. In light of Wormald’s theorem, we have done most of the work. Let D open contain the compact domain $[0, \tau] \times [0, 1]^K$. The stopping time T_D is defined to be the minimum t such that $(2t/n, \vec{R}(t)) \notin D$. We already have the boundedness condition with $\lambda_1 = O(K^2/n)$, as well as the trend condition. As $|\Omega| < \infty$,

each of the differential equation, f , is polynomial in ρ_i , which is C^∞ . Hence, the Lipschitz condition follows as the derivative of f is continuous on the closure of D .

Hence, by Wormald's theorem, we have convergence up to $0 \leq t \leq \sigma_n$, where σ represents the limits to which the solution to the differential equation can be extended. To finish the proof, by Theorem 24, the differential equation can be extended arbitrarily, and hence $\sigma = \tau$, as desired. $\square$

Remark 3. *The above proof, while important in its own right, is a showcase of the DEM technique. Much of the effort goes into massaging the quantity of interest into a difference equation with a small error. Moreover, it is deceptively difficult to analyze the differential equation: is the solution unique? How long does it extend?*

Thus, one can perhaps criticize the differential equation method for being less helpful than advertised: one has simply transferred the difficult questions of percolation and phase transition into a system of differential equations. However, ODEs are typically easier to analyze, and much more amenable to numerical simulation. Furthermore, the arguments made to prove Wormald's theorem are often applicable to proving uniform convergence of a stochastic process to a function ρ .

5.3.2 Susceptibility

More closely related to the global convergence of G_t is the susceptibility of the random graph G_t . Almost an identical argument using DEM can be given to track the evolution of the susceptibility of the random graph.

Definition 8. *The susceptibility of a graph $S(G)$ is the expected component size of a vertex selected uniformly at random. In other words, $S(G) = \frac{1}{n} \sum_{k=1}^{\infty} kN_k$.*

The susceptibility of the random graph and the size of its largest component clearly satisfy: $L_1^2/n \leq S(G) \leq L_1$. Hence, if a giant component of size $\Theta(n)$ emerges, $S(G) \geq \frac{L_1}{n} \cdot L_1 \mapsto \infty$. Hence, the emergence of the giant component clearly implies the divergence of the susceptibility. The converse is yet unknown to hold in general ([27]), although [29] demonstrates that in all bounded-size Achlioptas processes, the giant component emerges precisely when the susceptibility diverges. For those reasons, the blow-up point of the susceptibility is of great interest in random graph dynamics and can be analyzed via DEM.

As we are analyzing bounded-size Achlioptas processes, one can let $S_\omega(G) = S(G) - \sum_{i=1}^K iR_i(G) = \frac{1}{n} \sum_{|C_i| > K} |C_i|^2$ to be the essential susceptibility. Clearly, if WLOG $\{v_1, v_2\}$ is selected,

$$S^*(i+1) = S(i) + \frac{1}{n} 2|C(v_1)||C(v_2)| \quad (5.19)$$

as long as the round is not redundant. As before, we can define e_S to be the error term introduced in assuming that the round is non-redundant, and one can bound:

$$0 \geq E[e_S] \geq -\frac{4}{n^3} \sum |C_i|^4 \quad (5.20)$$

This bound is more troublesome than the case for R . For the sake of our purposes, let us ignore this error term, and move onto the differential equation. As for the trend equation, we can casework on the size of the components joined by the new edge to obtain a differential equation similar to the one obtained for R_k . Critically, the case in which two “large” components (components of size ω) are joined, the normalized S changes by a factor of:

$$\sum_{c(v_1)=c(v_2)=\omega} |C(v_1)| \cdot |C(v_2)| = n^2 S_\omega^2 \implies \Delta S = R_{j_3} R_{j_4} S_\omega^2 \quad (5.21)$$

This essentially shows that S'_ω has a positive coefficient of S_ω^2 in it, which implies loosely that S_ω and hence S , diverges.

As stated earlier, Spencer and Wormald ([29]) show that the blowup point of the differential equations governing S , denoted as t_b , is precisely t_c , the percolation point in the random graph process where the giant component emerges. Hence, the differential equation for S gives a single differential equation that allows us to numerically estimate the percolation point t_c for general bounded-size Achlioptas processes.

Chapter 6

Bounded-Size Achlioptas Processes: a Universality Class for Erdős-Rényi Phase Transition

We are now ready to prove the results of [25], the universality of Erdős-Rényi transition behavior for bounded-size Achlioptas processes. We shall closely follow the exposition given in [25]. The original paper is almost 100 pages long, and naturally we cannot give a full replication of the precise results of the paper. Hence, we shall simplify our exposition by giving a much simpler proof of a slightly weaker result. To begin, we shall give a brief summary of the results, which we already introduced in Chapter 3.

6.1 Statement of the Main Result

Simply put, we wish to prove that certain qualitative features of the phase transition for Erdős-Rényi graph processes extend to bounded-size Achlioptas processes. Recall that an Achlioptas process is “bounded-size” if it treats all components of order bigger than K “equally” in its edge selection.

What does it mean for bounded-size Achlioptas processes to exhibit an Erdős-Rényi-like phase transition behavior? It has both a local and global meaning. Locally, it means that the relative frequency of component sizes follow similar asymptotics to the Erdős-Rényi graph (equivalently, the fraction of vertices in components of fixed size exhibit similar asymptotics). Global convergence, on the other hand, focuses on the behavior of the largest component, and states that the size of the largest component behaves similarly to the Erdős-Rényi graph. In particular, the giant component also emerges in bounded-size Achlioptas processes, and its growth is initially analytic for a neighborhood of the critical time.

In both the local and global formulations, we study the behavior of normalized random graph quantities ρ and ρ_k . Recall that N_k is the number of vertices in the random graph in components of size k , and L_1 is the size of the largest component of the random graph. Then, the normalized quantities $\rho_k(t) = N_k(G_{tn})/n$ and $\rho(t) = L_1(G_{tn})/n$ are often easier to study. For example, we have seen in the DEM section that for bounded-size Achlioptas processes, ρ_k converges to a smooth function that follows a finite set of ordinary differential equations. While such a convenient differential equation does not exist for ρ , [26] proves for general Achlioptas processes that $\rho(t) = 1 - \sum_{k=1}^{\infty} \rho_k(t)$. In particular, we thus obtain that $L_1(G_{tn})/n$ converges to $\rho(t)$.

Hence, we can formulate our universality result using the normalized functions ρ and ρ_k . For visibility, we shall repeat the key theorems.

Theorem 26 (Local Asymptotics). *For bounded-size ℓ -vertex rule Achlioptas processes, assuming all possible component sizes k can be reached, we have for $t \in [t_c - \epsilon_0, t_c + \epsilon_0]$:*

$$\rho_k(t) = (1 + O(1/k))k^{-3/2}\theta(t)e^{-\psi(t)k} \quad (6.1)$$

Moreover, convergence of N_k to ρ_k is “uniform” in k and t near $t = t_c$: for i such that $\epsilon = i/n - t_c$ satisfies $10\psi''(t_c)\epsilon^2 k \leq \log n$, one simultaneously has for all such i and $1 \leq k \leq n^{1/10}$:

$$\begin{aligned}\frac{N_k(i)}{n} &= (1 \pm n^{-1/30}/k) \cdot \rho_k(i/n) \\ \frac{N_{\geq k}(i)}{n} &= (1 \pm n^{-1/30}/k) \cdot \left(\sum_{j \geq k} \rho_j(i/n) + \rho(i/n) \right)\end{aligned}\tag{6.2}$$

Theorem 27 (Global Asymptotics). *Let R be a bounded-size ℓ -vertex rule with critical time $t_c > 0$.*

1. *Subcritical case: for $i/n = t_c - \epsilon$, $\epsilon \in [0, \epsilon_0]$ with $\epsilon^3 n \mapsto \infty$, we have:*

$$L_r(i) = \psi(t_c - \epsilon)^{-1} \left(\log(\epsilon^3 n) - \frac{5}{2} \log \log(\epsilon^3 n) + O(1) \right)\tag{6.3}$$

where L_r is the size of the r -th largest component.

2. *Supercritical case: $L_1(tn)/n \mapsto \rho(t)$ in probability, where ρ is 0 for $t < t_c$, $\rho(t_c) = 0$, and $\rho(t)$ is analytic on $[t_c, t_c + \delta]$, with the right derivative of ρ at t_c strictly positive. In other words, for $\epsilon \in [0, \epsilon_0]$,*

$$\rho(t_c + \epsilon) = \sum_{j=1}^{\infty} a_j \epsilon^j\tag{6.4}$$

Furthermore, if $\epsilon = i/n - t_c$ satisfies $\epsilon^3 n \geq \omega$, $\omega \mapsto \infty$, $\tau = (\log \omega)^{-1/2}$ we have concentration around ρ , uniformly for all i :

$$L_1(i) = (1 \pm \tau) \rho(t_c + \epsilon) n, \quad L_2(i) \leq \tau L_1(i)\tag{6.5}$$

It can be helpful to compare the results with the original results for Erdős-Rényi processes, which we summarize below. For a more detailed discussion on the comparison, see Chapter 3.

– Erdős-Rényi Process

• Distribution of small components near $t_c = 1/2$:

$$\rho_k(1/2 + \epsilon) \approx \frac{1}{\sqrt{2\pi}} k^{-3/2} e^{-k(2+o(1))\epsilon^2}\tag{6.6}$$

• Largest component, subcritical period: for $t < t_c$, the components are uniformly small:

$$L_r(t) = \frac{1}{\alpha} (\log n - \frac{5}{2} \log \log n + O(1))\tag{6.7}$$

• Largest component, supercritical period:

$$\rho(1/2 + \epsilon) = 4\epsilon + \sum_{j=2}^{\infty} c_j \epsilon^j\tag{6.8}$$

– Bounded-size Achlioptas Processes

• Distribution of small components near $t = t_c$:

$$\rho_k(t) = (1 + O(1/k)) k^{-3/2} \theta(t) e^{-\psi(t)k}\tag{6.9}$$

• Largest component, subcritical period: $t = t_c - \epsilon$:

$$L_r(t) = \psi(t_c - \epsilon)^{-1} \left(\log(\epsilon^3 n) - \frac{5}{2} \log \log(\epsilon^3 n) + O(1) \right)\tag{6.10}$$

• Largest component, supercritical period:

$$\rho(t_c + \epsilon) = \sum_{j=1}^{\infty} a_j \epsilon^j\tag{6.11}$$

Theorems 26 and 27 make precise the notion that bounded-size Achlioptas processes constitute a universality class for the phase-transition behavior of Erdős-Rényi processes: the normalized size of the giant component follows a continuous function which is analytic beyond criticality, and the normalized component distribution is asymptotically proportional to $k^{-3/2} e^{-\theta k}$.

6.1.1 Simplification of the Result

The essence of the proof in [25] stems from coupling the component size distribution of the Achlioptas process with a branching process family $\mathcal{S}_t$. While this basic picture of the proof is intuitive and illustrative, much of the paper involves heavily technical computation bounding the second moments of the number of vertices in large components.

For our expositional purposes, the argument can be simplified significantly if we prove a slightly weaker result. We shall still be able to prove the local asymptotics theorem, Theorem 26, but we shall prove a simpler version of the global asymptotics theorem, Theorem 27.

There are two main simplifications: first, for the subcritical case, we shall only prove a weak upper bound for L_1 , namely the fact that $L_1(i)$ is at most $o(n)$, such that the normalized function $L_1(i/n)/n = 0$ in probability for the subcritical case. Second, instead of proving the uniform convergence of $L_1(i)$ to ρ , we shall instead prove pointwise convergence. Intuitively, proving this weaker form of convergence will save us from needing the precise concentration/second moment bounds obtained via strenuous work in the original paper.

To summarize, the following is the simplified theorem that we shall prove.

Theorem 28 (Global Asymptotics, Simplified). *Let R be a bounded-size ℓ -vertex rule with critical time $t_c > 0$.*

1. *Subcritical case: for $i/n = t_c - \epsilon$, $\epsilon \in [0, \epsilon_0]$ with $\epsilon^3 n \mapsto \infty$, we have:*

$$L_1(i) \leq O(\log n^D n^{1/2}) = o(n) \quad (6.12)$$

2. *Supercritical case: $L_1(tn)/n \mapsto \rho(t)$ in probability, where ρ is 0 for $t < t_c$, $\rho(t_c) = 0$, and $\rho(t)$ is analytic on $[t_c, t_c + \delta]$, with the right derivative of ρ at t_c strictly positive. In other words, for $\epsilon \in [0, \epsilon_0]$,*

$$\rho(t_c + \epsilon) = \sum_{j=1}^{\infty} a_j \epsilon^j \quad (6.13)$$

For the rest of this chapter, we shall simultaneously prove Theorem 26 and 28. The set-up of the proof requires an elaborate and ingenious two-round exposure argument, which we shall illustrate in the next section.

6.2 Preliminaries: Two-round Exposure, The Poissonized Graph, and the Concentration of the Parameter List

6.2.1 Two-round Exposure, the Auxillary Graph H_i , and the parameter list $\mathcal{G}_i$

Both Theorems 26 and 27 zoom in on the Achlioptas process near $t = t_c$. Hence, we shall consider $G(i)$ from $i_0 = (t_c - \sigma)n$ to $i_1 = (t_c + \sigma)n$, with the constant σ to be chosen opportunely. The proof relies on the following two-round exposure argument, which exploits the bounded-size property of the Achlioptas process. Intuitively, by revealing the information in two rounds, one can reintroduce i.i.d uniform edge selection, which enables branching process methods.

Aside: a quick parity assumption

The original paper notes the possibility that certain component sizes can never be reached. For example, if a rule never joins an isolated vertex to any component other than an isolated vertex, it is easy to see that the component sizes are all even.

While this is an interesting complication, for our purposes, we shall focus on the case where all component sizes are reachable. The original 2-edge Achlioptas process satisfies this: if both selected edges connect a

component of size k to size 1, it must create a component of size $k + 1$. (In fact, one can show that the set of reachable component sizes are multiples of powers of 2, hence the term “parity” assumption).

One important, and intuitive, consequence of our parity assumption is that $\rho_k(t) > 0$ for all k , i.e. there is asymptotically a positive proportion of vertices in all component sizes. To sketch the argument, the proof follows via induction on k . Intuitively, given that all k is reachable, there is an ℓ -tuple $(c_1, c_2, \dots, c_\ell)$ that will yield k . The probability that such an ℓ -tuple will be selected is asymptotically bounded above 0 by induction ($\rho_{c_i} > 0$). Hence, k will appear with asymptotically significant positive probability, and will stay long enough (i.e. if none of the components get selected).

Two-round Exposure

Suppose we are given $G(i_0)$, which is at its subcritical stage. One can partition the vertices of $G(i_0)$ into those belonging to components of size K or less, and those belonging to components of size greater than K . We shall denote the former as V_S and the latter as V_L . Needless to say, $|V_S| + |V_L| = n$. To reiterate, V_S and V_L depend *only* on the graph $G(i_0)$: a vertex $v \in V_S$ may belong to a component of size greater than K in $G(i)$ as more edges are added.

We shall reveal the random graph process from $i = i_0$ to $i = i_1$ in two rounds. Naturally, if one reveals the $(i_1 - i_0)$ ℓ -tuples of vertices for each step, one is able to fully determine the random graph.

In the first round, for each ℓ -tuple corresponding to $i_0 \leq i \leq i_1$, we shall only reveal the vertices in V_S , not revealing the identity of the remaining V_L vertices. The second round exposure simply reveals the identity of the V_L vertices.

Lemma 6. *The information revealed by the first exposure round is enough to make all decisions by the rule R (i.e. which edges out of the ℓ -tuple to select). Furthermore, the vertices revealed in the second exposure round is equivalent to filling in the remaining V_L vertices of the first round independently and uniformly random from V_L .*

Proof. As $\vec{v}_i$ are chosen independently and uniformly at random, the statement about the second-round follows immediately.

The statement about the first-round follows from the fact that R is a bounded-size rule and the observation that the component size $\vec{c}_i \in \{1, \dots, K, \omega\}^\ell$ is determined from the first-exposure round. One can show this via induction. In the inductive step, if one adds a $V_S - V_S$ edge, we can precisely update the component size of each vertex. If one adds a $V_S - V_L$ vertex, the vertices in the same component as the V_S vertex now belong to a component of size ω . A similar argument holds for $V_L - V_L$ edges. Hence, one precisely knows the component size for every ℓ -tuple from the first exposure round, which is all that the decision rule needs. $\square$

By the lemma, after the first exposure round, we know exactly which $V_S - V_S$ edges were added, and for each V_S vertex v_S , we know how many edges connect v_S to a v_L vertex (called “stubs”), as well as how many $V_L - V_L$ edges are added.

Auxiliary Graph H_i and the Parameter List $\mathcal{G}_i$

Starting from the original graph $G(i_0)$, one can obtain the auxiliary graph H_i for $i_0 \leq i \leq i_1$ as follows:

1. Insert all $V_S - V_S$ edges added in steps i_0 to i .
2. Attach, for each $v_S - V_L$ edge added in steps i_0 to i , “stubs” connected to v_S , i.e. half-edges.
3. Keep track of how many $V_L - V_L$ edges are added from i_0 to i .

As we only care about component sizes, the internal structure of each component is irrelevant. Also, note that the components in V_L in $G(i_0)$ remain unchanged in H_i , whereas some of the components in V_S

may merge. A V_S component of H_i is of type (k, r) if it is of size k and has r stubs emanating out of it. Let $Q_{k,r}(i)$ be the number of components of type (k, r) of H_i . Naturally, the following hold:

$$Q_{k,r}(i_0) = N_k(i_0)/k, |V_L| = \sum_{k>K} N_k(i_0), |V_S| = \sum_{k\geq 1} kQ_{k,r}(i) \quad (6.14)$$

It will be helpful to think of each $V_L - V_L$ edge as a component of type $(0, 2)$: it connects an “empty” component in V_S to two stubs to V_L . Hence, we shall denote $Q_{0,2}(i)$ as the number of $V_L - V_L$ edges added up to time i , with $Q_{0,r}(i) = 0$ for $r \neq 2$.

The parameter list of H_i is simply the size of the components in V_L as well as the number of components of type (k, r) . In other words,

$$\mathcal{G}_i = ((N_k(i_0))_{k>K}, (Q_{k,r}(i))_{k,r}) \quad (6.15)$$

$\mathcal{G}_i$ contains all the relevant information of H_i , i.e. it serves as a type of sufficient statistics for the component size distribution of G_i . Conditioned on $\mathcal{G}_i$, one can obtain a random graph $J(\mathcal{G}_i)$ by connecting each stub to a uniformly random vertex in V_L , and then add $Q_{0,2}$ random edges in V_L .

By construction and the previous lemma, we obtain that the random graph $J(\mathcal{G}_i)$ has the same component size distribution as G_i , conditioned on the parameter list $\mathcal{G}_i$. Hence, we have reduced the study of the component distribution of $\mathcal{G}_i$ to the study of the randomized graphs $J(\mathcal{G}_i)$.

The Poissonized random graph J^{Po} and the hyperedge formulation

We shall need yet another reduction from $J(\mathcal{G}_i)$.

Definition 9. Let $\mathcal{G} = ((N_k)_{k>K}, (Q_{k,r}))$ be a parameter list with $N_k \in \mathbb{N}$ and $Q_{k,r} \in [0, \infty)$. The Poissonized random graph $J^{Po}(\mathcal{G})$ is defined in the same way as $J(\mathcal{G})$, except that the numbers of (k, r) -type components are now independent Poisson random variables with mean $Q_{k,r}$.

Now, we shall think about each (k, r) component as a “weighted-hyperedge” on vertices in V_L . In other words, when one connects a (k, r) component to r -tuple of random vertices in V_L , $(v_1, v_2, \dots, v_r)$, one can instead view it as adding a hyperedge of weight k to $(v_1, v_2, \dots, v_r)$. Needless to say, the $(0, 2)$ components are simple random edge added to V_L .

The hyperedge formulation allows us to focus on $H_L = H_i[V_L]$, i.e. the auxiliary graph H_i restricted to V_L . H_L then has N_k/k components of size $k > K$, to which we add $Q_{k,r}$ hyperedges of weight k .

By standard splitting results in Poisson processes, the Poissonization J^{Po} then allows us to consider each hyperedge $g = (v_1, \dots, v_r) \in (V_L)^r$ of weight k separately as independent Poisson processes with rate:

$$\mu_{k,r} = Q_{k,r}/|V_L(\mathcal{G})|^r \quad (6.16)$$

Thus, for each vertex $v_L \in H_L$, one can consider all r -tuples containing v_L , and sum the independent Poisson random variables to obtain all the hyperedges containing v_L . Looking ahead, this will allow us to estimate the neighborhood exploration process with a related branching process, where the offspring distribution is “almost” a sum of independent Poisson r.v.s.

The Poissonized graph J^{Po} , modeled via the addition of independent (k, r) hyperedges to H_L , will be the focus of our analysis.

To summarize, we have reduced the analysis of the component distribution of G_i into the following steps:

$$G_i \xrightarrow{\text{First-round Exposure}} H_i \xrightarrow{\text{Sufficient Statistics}} \mathcal{G}_i \xrightarrow{\text{Connecting random } V_L \text{ vertices}} J(\mathcal{G}_i) \xrightarrow{\text{Poissonization}} J^{Po}(\mathcal{G}_i)$$

We can then analyze $J^{Po}(\mathcal{G})$ via a branching process argument. Of course, J^{Po} is not equivalent to J , and some argument is required to relate J^{Po} to J .

6.2.2 Evolution of $\mathcal{G}_i$: Applications of DEM

In the previous section, we have roughly reduced the study of the component distribution of G_i into a study of the Poissonized random graph $J(\mathcal{G})$. However, we still have not made any claims about the parameter list $\mathcal{G}$, which is revealed after the first-round exposure.

I claim that $\mathcal{G}$ is concentrated around a deterministic function, and is furthermore subexponential. Specifically, we can define $\mathcal{G}$ to be t -nice as follows:

Definition 10 (t -nice parameter lists). *For $t \in [t_0, t_1]$, a parameter list $\mathcal{G}$ is t -nice if it satisfies the following conditions:*

1. *The quantities N_k and $Q_{k,r}$ are concentrated: there exists continuous (in fact, analytic) functions ρ_k and $q_{k,r}$ such that*

$$\begin{aligned} |N_k - \rho_k(t_0)n| &\leq (\log n)^{D_N} n^{1/2} \text{ for all } k > K \\ |Q_{k,r} - q_{k,r}(t)n| &\leq (\log n)^{D_Q} n^{1/2} \text{ for all } k, r \geq 0 \end{aligned} \quad (6.17)$$

2. *The quantities N_k and $Q_{k,r}$ are subexponential: there exists constants A, a and B, b such that*

$$\begin{aligned} N_{\geq k} &\leq Ae^{-ak}n \text{ for all } k > K \\ Q_{k,r} &\leq Be^{-b(k+r)}n \text{ for all } k, r \geq 0 \end{aligned} \quad (6.18)$$

As the naming suggests, the parameter list of G_i is “nice” with high-probability:

Theorem 29 (Parameter lists of G_i are uniformly nice with high probability). *Let N_i denote the event that the random parameter list $\mathcal{G}_i = \mathcal{G}(G_{n,i})$ is t -nice, where $t = i/n$. Let $N = \cap_{i_0 \leq i \leq i_1} N_i$. Then,*

$$P(N) = 1 - O(n^{-99}) \quad (6.19)$$

In this section, we shall give a proof of the concentration result. The subexponentiality results are derived in [27] after extensive analysis, and in the interest of expositional clarity, we shall only give a sketch of the proof. Together, they yield the proof of Theorem 29.

The essential technique in proving the concentration of N_k is the differential equation method (DEM). In fact, we have already applied the technique to bounded-size Achlioptas processes and obtained the concentration result: see the local convergence section of Chapter 5. Let us repeat the result below:

Concentration of N_k

Theorem 30. *Let $C = \{1, 2, \dots, K, \omega\}$, and N_k be the number of vertices in components of size k . With probability at least $1 - n^{-\omega(1)}$, we have:*

$$\max_{0 \leq i \leq i_1} \max_{k \in C} |N_k(i) - \rho_k(i/n)n| \leq (\log n)n^{1/2} \quad (6.20)$$

with $\sum_{k \in C} \rho_k(t) = 1$, $\rho'_\omega(t) \geq 0$, $\Delta_\rho^R \leq 2K$, and

$$\rho'_k(t) = \sum_{\vec{c}} \Delta_\rho^R(k, \vec{c}) \prod \rho_{c_j}(t) \quad (6.21)$$

Proof. Let $\Delta_\rho^R(k, \vec{c})$ be the deterministic change in the number of vertices in components of size k when $\vec{c}$ is the vector of component sizes of the selected ℓ -tuple, and each vertex belongs to a distinct component. Clearly, $\Delta_\rho^R \leq 2K$. We need to show that two main conditions to use the DEM are satisfied:

Bounded Differences: $|N_k(i+1) - N_k(i)| \leq 2K$

$$\text{Difference Equation: } E(N_k(i+1) - N_k(i) | G_i) = \sum_{\vec{c} \in C^\ell} \left(\Delta_\rho^R(k, \vec{c}) \prod_{j \in [\ell]} \frac{N_{c_j}(i)}{n} \right) \pm \frac{4\ell^2 K^2}{n} \quad (6.22)$$

The first identity is obvious from the definitions. To prove the difference equation, as long as there are no small (smaller than $K + 1$) components chosen multiple times in the ℓ -tuple, N_k evolves precisely according to Δ_ρ^R . Furthermore, the probability that at least two of the ℓ randomly chosen vertices lie in the same component of size at most K is at most $\ell^2 K/n$. As the difference $|N_k(i+1) - N_k(i)| \leq 2K$, the difference equation follows.

Finally, it suffices to show that the differential equation 6.21 extends to $t = [0, \infty)$. This we have done in the DEM section, by showing that as $\rho_k \in [0, 1]$, one obtains absolute bounds on ρ'_k , which are given by polynomial equations of ρ . Basic results in ODE then imply that these solutions can be extended arbitrarily for all time. For our purposes, we only need the extension to $[t_0, t_1]$. (This automatically implies the Lipschitz condition)

Then, we can use Wormald's theorem to obtain the result, plugging in $\gamma = \log nn^{1/2}$. As for basic properties of ρ_k , one can recursively differentiate the ODE to obtain that ρ_k is smooth, and as $\sum N_k(i) = n$, $\sum_{k \in C} \rho_k = 1$. Finally, one can see easily that $\Delta_\rho^R(\omega, \vec{c}) \geq 0$, and hence $\rho'_\omega(t) \geq 0$. $\square$

For our purposes, we need to extend the above result to an arbitrary $k' \geq K$ (provided k' is fixed).

Corollary 10. *Fixing k' , with probability at least $1 - n^{-\omega(1)}$, we have*

$$\begin{aligned} \max_{0 \leq i \leq i_1} \max_{1 \leq k \leq k'} |N_k(i) - \rho_k(i/n)n| &\leq (\log n)n^{1/2} \\ \max_{0 \leq i \leq i_1} \max_{1 \leq k \leq k'} |N_{\geq k}(i) - \rho_{\geq k}(i/n)n| &\leq (\log n)n^{1/2} \end{aligned} \quad (6.23)$$

where $\rho_k : [0, t_1] \mapsto [0, 1]$ are given by an extension of the same differential equations for $k \leq K$, with $\rho_{\geq k}(t) = 1 - \sum_{1 \leq j < k} \rho_j(t)$.

Proof. For $k \leq K$, the equations above remain valid, and thus no need to repeat the proof. For $k > K$, one can repeat the argument given above, except we shall let the possible component sizes be $\{1, 2, \dots, K, \dots, k', \omega\}$. Then, one can define $\Delta_\rho^R(k, \vec{c})$ in the same way, and note that each ρ'_k depends only on ρ_j with $j \leq \max\{k, K\}$. A repetition of the above argument then gives the theorem. $\square$

A cursory examination of the differential equations above yields that $|V_S|$ and $|V_L|$ are $\Theta(n)$, a fact we shall need later. It suffices to show that both ρ_1 and ρ_ω are strictly positive for $t \in (0, t_c + \sigma]$.

Lemma 7. *For all $t \in (0, t_c + \sigma]$, we have $\min\{\rho_1(t), \rho_\omega(t)\} > 0$*

Proof. First, note that $\Delta_\rho^R(k, \vec{c}) \geq 0$ if $k \notin \vec{c}$. Secondly, as all $0 \leq \rho_j(t) \leq 1$ and $\sum \rho_j = 1$, with $|\Delta| \leq 2k$, it follows for any integer $k \geq 1$:

$$\rho'_k(t) \geq -2\ell k \rho_k(t) \implies (\rho_k e^{2\ell k t})' \geq 0 \quad (6.24)$$

Now, I claim that $\rho_{2^j}(t) > 0$ for $t \in (0, t - 1]$, inducting on $j = 0$. For ρ_1 , as $\rho_1(t)e^{2\ell t}$ is increasing, $\rho_1(t) \geq e^{-2\ell t} > 0$, as desired. To induct, note that $\Delta_\rho^R(k, (k/2, k/2, \dots, k/2)) = k \geq 1$. Hence, $\rho'_{2^j}(t) \geq (\rho_{2^{j-1}}(t))^\ell - 2 \cdot 2^j \ell \rho_{2^j}(t) \implies (\rho_{2^j}(t)e^{2\ell 2^j t})' \geq (\rho_{2^{j-1}}(t)e^{2\ell t})^\ell$. As $\rho_{2^{j-1}}(t) > 0$ by induction, on the compact interval there is a δ such that $\rho_{2^{j-1}}(t')e^{2\ell t'} > \delta$ on $t' \in [t/2, t]$. This implies $(\rho_{2^j}(t)e^{2\ell 2^j t})' > \delta^\ell$ in $[t/2, t]$, and hence $\rho_{2^j}(t) > 0$, as desired. We conclude by noting $\rho_\omega > \rho_{2^j}$ for $2^j > K$. $\square$

Concentration of $Q_{k,r}$

Next, we prove the concentration of $Q_{k,r}$, also using the differential equation method.

First, let us track the evolution of $Q_{0,2}$, the number of $V_L - V_L$ edges. Note that $Q_{0,2}$ only changes when we select two vertices in V_L , so we label the vertices in V_L as L (this is treated the same way as ω by R , thanks to boundedness). Again, provided that all of the vertices we select are from different components, the number of V_L edges increase by a deterministic quantity $\Delta_{0,2}^R(\vec{c})$.

In this case, we need to treat vertices in components of size ω in V_S differently from the V_L vertices. Hence, let $\theta_L(t) = \rho_\omega(t_0)$ be the idealized size of V_L , and let $\theta_k(t)$ be $\rho_k(t)$ for $k \leq K$, and $\theta_\omega(t) = \rho_\omega(t) - \rho_\omega(t_0)$, which stands for V_S vertices in components of size ω .

Thus, now setting $C' = \{1, 2, \dots, K, \omega\} \cup \{L\}$, we obtain that $Q_{0,2}(i)$ concentrates around $q_{0,2}(t)$, which is the unique solution to the differential equation:

$$q_{0,2}(t_0) = 0, \quad q'_{0,2}(t) = \sum_{\vec{c} \in C'} \Delta_{0,2}^R(\vec{c}) \prod_{j \in [\ell]} \theta_{c_j}(t) \quad (6.25)$$

We shall omit the precise justification for Wormald's theorem (bounded-difference is trivial, the difference equation can be justified in much of the same way as we had for N_k , and Lipschitz follows from the well-behavedness of the differential equation), and obtain:

Theorem 31. *With probability at least $1 - n^{-\omega(1)}$ we have:*

$$\max_{i_0 \leq i \leq i_1} |Q_{0,2}(i) - q_{0,2}(i/n)n| \leq (\log n)^2 n^{1/2} \quad (6.26)$$

with $q'_{0,2}(t) > 0$ (as all the Δ are positive.)

Finally, we track $Q_{k,r}$, the components in V_S of size k and r stubs. To obtain the precise differential equations, note that a (k, r) -type component is created under two circumstances. First, a $V_S - V_S$ type edge is added and joins two components of size k_1, k_2 with stubs r_1, r_2 , such that $k_1 + k_2 = k$, and $r_1 + r_2 = r$. Second, a $V_S - V_L$ edge is added to a $(k, r-1)$ -type component. Finally, note that the (k, r) -type component is destroyed iff one adds an edge to a preexisting (k, r) component.

To simplify notation, let $s(k, r) = \omega$ if $k \geq K + 1$ or $r \geq 1$, and k otherwise, and let the Achlioptas rule R connect v_{j_1}, v_{j_2} . The discussion above yields the following quantities:

$$\begin{aligned} V_S - V_S \text{ edge: } A(k, r, \vec{c}) &= \sum_{k_1+k_2=k: k_i \geq 1, r_1+r_2=r} k_1 q_{k_1, r_1}(t) k_2 q_{k_2, r_2}(t) I(c_{j_1} = s(k_1, r_1), c_{j_2} = s(k_2, r_2)) \\ V_S - V_L \text{ edge: } B(k, r, \vec{c}) &= I(r \geq 1) k q_{k, r-1} \rho_\omega(t_0) (I(c_{j_1} = s(k, r-1), c_{j_2} = \omega) + I(c_{j_2} = s(k, r-1), c_{j_1} = \omega)) \\ \text{Destruction: } C(k, r, \vec{c}) &= k q_{k, r} (I(c_{j_1} = s(k, r)) \rho_{c_{j_2}}(t) + I(c_{j_2} = s(k, r)) \rho_{c_{j_1}}(t)) \end{aligned} \quad (6.27)$$

Thus, one can express:

$$q'_{k,r} = \sum_{\vec{c}} \left[\prod_{j \neq \{j_1, j_2\}} \rho_{c_j}(t) \right] (A(k, r, \vec{c}) + B(k, r, \vec{c}) - C(k, r, \vec{c})) \quad (6.28)$$

The key takeaway from this complicated differential equation is that $q'_{k,r}$ only depend on $q_{k',r'}$ with $k' \leq k$ and $r' \leq r$, and $(\rho_j)_{j \in C}$, i.e. only finitely many quantities. Hence, standard results in ODE imply that the infinite system of differential equations has a unique solution, and smoothness can be obtained by repeatedly differentiating the differential equations. Furthermore, $0 < q < 1$ from the definition.

Hence, we obtain:

Theorem 32. *Given $k' \geq 1$ and $r' \geq 0$, with probability at least $1 - n^{-\omega(1)}$ we have:*

$$\max_{i_0 \leq i \leq i_1} \max_{1 \leq k \leq k', 0 \leq r \leq r'} |Q_{k,r}(i) - q_{k,r}(i/n)n| \leq (\log n)^2 n^{1/2} \quad (6.29)$$

It is important to note that the above concentration result is not for all k and r , but only an arbitrarily finite amount. To extend concentration to all k and r , one needs the subexponentiality result of the next section.

Subexponentiality of N_k and $Q_{k,r}$

While we have established the concentration of N_k and $Q_{k,r}$ to solutions of certain differential equations, we need for our purposes strong (subexponential) bounds on $Q_{k,r}$. This relies on ideas established in [27], a previous paper by the same co-authors. Giving a precise account of their argument is beyond the scope of this exposition, so we shall give a rough sketch of their argument.

First, for $N_k(i_0)$, we exploit that the random graph is subcritical, and use a different two-round exposure argument. For a linear number of steps (δn), we expose first the total ℓ -tuples presented to R , and then in the second round reveal the order of the ℓ -tuples. Then, the first round severely restricts the number of components and tuples that can change the size of the component containing v . In particular, the only relevant component/tuples are those that can be reached from v by adding all $\binom{\ell}{2}$ edges of all the ℓ -tuples appearing in the first exposure phase.

Then, one can crudely upper-bound the total number of offspring distribution for this exploration process: as there are at most $\ell n^{\ell-1}$ ℓ -tuples containing v , with each introducing at most $(\ell - 1)$ new vertices, the expected size of the offspring after δn steps is bounded above by:

$$\delta n \cdot \frac{\ell n^{\ell-1}}{n^\ell} \cdot (\ell - 1) \cdot \sum_{k \geq 1} k N_k(G)/n = \delta \ell (\ell - 1) S(G) \quad (6.30)$$

where $S(G)$ is the susceptibility of G . Hence, if δ is such that $\delta \ell (\ell - 1) S(G) < 1$, the above “idealized” exploration process is subcritical, which establishes tight concentration and exponential bounds. We can find such a δ as $S(G) < \infty$, as the random graph process is subcritical.

For $Q_{k,r}$, as we need to go beyond $t = t_c$, we need to tweak the argument above. We can achieve that by marking all the V_L neighbors as belonging to components of size $K + 1$ (or more precisely, the exploration process halts if it hits a V_L vertices).

The heuristic argument above yields the following theorem:

Theorem 33 (Subexponentiality of the Parameter List). *There exists constants A, a, B, b , and D such that with probability $1 - n^{-99}$ the following hold for all k, r :*

$$\begin{aligned} \max_{0 \leq i \leq i_0} |N_k(i) - \rho_k(i/n)n| &\leq (\log n)^D n^{1/2} & \max_{i_0 \leq i \leq i_1} |Q_{k,r}(i) - q_{k,r}(i/n)n| &\leq (\log n)^D n^{1/2} \\ \max_{0 \leq i \leq i_0} N_{\geq k}(i) &\leq A e^{-ak} n & \max_{i_0 \leq i \leq i_1} Q_{\geq k, \geq r}(i) &\leq B e^{-b(k+r)} n \\ \sup_{t \in [0, t_0]} \rho_k(t) &\leq A e^{-ak} & \sup_{t \in [0, t_0]} q_{k,r}(t) &\leq B e^{-b(k+r)} \end{aligned} \quad (6.31) \quad (6.32)$$

Hence, the differential equation method, along with the branching process argument give us that with high probability the parameter list $\mathcal{G}$ is t -nice for all $i \in [i_0, i_1]$. Here are some quick and important consequences of $\mathcal{G}$ being uniformly nice.

Lemma 8. *The following hold:*

1. *There is an absolute constant $C > 0$ such that every nice parameter list satisfies $Cn \leq |V_L| \leq n$*
2. *There is a B_0 such that $Q_{k,r} = 0$ whenever $k + r \geq B_0 \log n$*
3. *Denoting $\alpha = (\log n)^2$, any nice parameter list satisfies*

$$\max_{k \geq \alpha} N_k = 0, \quad \max_{k+r \geq \alpha} Q_{k,r} = 0 \quad \text{and} \quad \max_{k,r} Q_{k,r} \leq \alpha |V_L| \quad (6.33)$$

Proof. The first is an obvious consequence of there being $k > K$ such that $\rho_k(t) > 0$, which follows from our parity assumption. The remaining identities follow from the exponential bound and the fact that $Q_{k,r}$ and N_k are integers. $\square$

6.3 Component size distribution: coupling arguments

For the rest of this chapter, we shall thus look at parameter lists $\mathcal{G} = ((N_k)_{k>K}, (Q_{k,r})_{k,r \geq 0})$ that are nice, and analyze the Poissonized graphs $J^{Po}(\mathcal{G})$ and $J(\mathcal{G})$.

The first step is to relate $EN_j(J^{Po}(\mathcal{G}))$ (in Poissonized graphs) to the associated vertex exploration process $T(\mathcal{G})$, also defined on the Poissonized graphs. Next, we shall relate the exploration process T to a sesqui-type branching process, which we already analyzed in Chapter 4. Intuitively, the fertile L -particles correspond to the neighbors in V_L , while the infertile S -particles correspond to the neighbors in V_S .

To summarize, we proceed in the following steps:

$$EN_j(\mathcal{G}) \xrightarrow{\text{Theorem 34}} |T(\mathcal{G})| \xrightarrow{\text{Coupling: Theorem 35}} P(|\mathcal{S}_t| = j)$$

6.3.1 Neighborhood Exploration Process: From component size distribution N_j to the Associated Random Walk T

We shall now describe the neighbourhood exploration process. Denoting $C_v(G)$ to be the component of G containing v , we wish to explore, from the initial set W , $C_W(J^{Po}) = \cup_{v \in W} C_v(J^{Po})$. As described earlier, we shall model the exploration process as an exploration on H_L , the restriction of J^{Po} to V_L , with the V_S vertices modeled as a hyperedge of weight k .

Then, let A_j and E_j be the active and explored V_L vertices after j steps of the exploration process. Furthermore, let S_j be the number of V_S neighbors encountered so far. We initialize with $E_0 = \emptyset$, $A_0 = W$, and $S = S_0$.

Given an active vertex $v_{j+1} \in A_j$, we sequentially test the presence and multiplicity of each untested (k, r) -hyperedge g containing v_{j+1} . Denote the “newly found” hyperedges as $\mathcal{H}_{j,k,r}$, and add to the active set $\cup_{1 \leq h \leq r} C_{w_h}(H_L) \setminus (A_j \cup E_j)$, i.e. the V_L neighbors of the newly introduced vertices of the hyperedge. Furthermore, for each $g \in \mathcal{H}_{j,k,r}$, we add k to S_j . Finally, we remove v_j from the active set to the explored set. We will stop when $|A_j| = 0$.

Let $M_j = |A_j \cup E_j|$ be the number of V_L vertices discovered by the exploration process up to step j . We wish to track $T = (M_j, S_j)_{j \geq 0}$, and denote $|T| = |C_W(J^{Po})| + S_0$, the total number of vertices reached.

Let $|\mathcal{G}| = \sum_{k>K} N_k + \sum_{k \geq 1, r \geq 0} kQ_{k,r}$ be the expected size of $J^{Po}(\mathcal{G})$ (for Poissonized graphs, the graph size is not deterministic). Now, how will we generate the initial W and S_0 ? It will depend on 2 cases: whether we start our exploration from $v \in V_L$ or $v \in V_S$.

1. In the first case, for each $v \in V_L$, we start the process with probability $\frac{1}{|\mathcal{G}|}$, with $S_0 = 0$ and $W = C_v(H_L)$.
2. In the second case, for each $k \geq 1, r \geq 0$, with probability $kQ_{k,r}/|\mathcal{G}|$ select the hyperedge $(w_1, \dots, w_r)$, and start with $S_0 = k$ and $W = \cup C_{w_h}(H_L)$.

Given such an initialization of W and S_0 , we obtain that the distribution of the associated random walk $|T|$ will be precisely the expected component size distribution of J^{Po} .

Theorem 34. *Let $|\mathcal{G}| = \sum_{k>K} N_k + \sum_{k \geq 1, r \geq 0} kQ_{k,r}$ be the expected size of $J^{Po}(\mathcal{G})$. Then, the expected component size distribution is given by the distribution of the random walk $|T|$.*

$$\begin{aligned} EN_j(J^{Po}) &= P(|T| = j)|\mathcal{G}| \\ EN_{\geq j}(J^{Po}) &= P(|T| \geq j)|\mathcal{G}| \end{aligned} \tag{6.34}$$

Proof. We shall need the following rather well known lemma in Poisson processes.

Lemma 9. *Let $N \sim Po(\lambda)$. Suppose that f is a random function that takes in finite sets and satisfies the following symmetry condition:*

$$E[f(x_1, x_2, \dots, x_n)] = E[f(x_{\sigma(1)}, \dots, x_{\sigma(n)})] \tag{6.35}$$

for any deterministic permutation of n (i.e. it is symmetric with respect to relabelling in expectations).
Then, the following is satisfied:

$$E\left[\sum_{i=1}^N f(x_i, X_N \setminus \{x_i\})\right] = \lambda E[f(x', X_N)] \quad (6.36)$$

where $X_N = \{x_1, \dots, x_N\}$, and x_i, x' are drawn independently from some common distribution.

Proof. Note by the Poisson pmf that $P(N = n) \cdot n = \lambda P(N = n - 1)$. Now, from the assumed independence, we have that the LHS above is:

$$\begin{aligned} E\left[\sum_{i=1}^N f(x_i, X_N \setminus \{x_i\})\right] &= \sum_{n \geq 1} P(N = n) E\left[\sum_{i=1}^n f(x_i, X_n \setminus \{x_i\})\right] = \sum_{n \geq 1} P(N = n) \cdot n \cdot E[f(x_1, x_2, \dots, x_n)] \\ &= \lambda \sum_{n \geq 0} P(N = n) E[f(x_1, x_2, \dots, x_{n+1})] = \lambda E[f(x', X_N)] \end{aligned} \quad (6.37)$$

where the first equality follows from symmetry, as desired. $\square$

Now, let us explicitly define the random function $f_H(g_1, g_2, \dots, g_n)$. Let g_i be a randomly selected (k, r) hyperedge, and $f_H(g, \{g_1, g_2, \dots, g_t\})$ be the size of the component containing the hyperedge g , where we add the (k, r) -hyperedges $\{g, g_1, g_2, \dots, g_t\}$ to H . The function is random as we still have random choices for other (k', r') hyperedges to add to H .

Next, let $\tilde{Q}_{k,r}$ be the set of (k, r) hyperedges selected for J^{Po} . The lemma gives us:

$$E\left(\sum_{g \in \tilde{Q}_{k,r}} I(|C_g(J^{Po})| = j)\right) = Q_{k,r} P(|C(J_{k,r}^{Po})| = j) \quad (6.38)$$

where $J_{k,r}^{Po}$ is obtained from J^{Po} by adding another k, r hyperedge.

Hence, we obtain an expression for $EN_j(J^{Po})$:

$$EN_j(J^{Po}) = \sum_{v \in V_L} P(|C_v(J^{Po})| = j) + \sum_{k \geq 1, r \geq 0} k Q_{k,r} P(|C(J_{k,r}^{Po})| = j) \quad (6.39)$$

Now, if we start our process at $v \in V_L$, $|T| = |C_v(J^{Po})|$, whereas starting the process at $v \in V_S$ by selecting a hyperedge with $S_0 = k$ and $W = \cup C_{w_h}(H_L)$ gives us $|T| = |C_W(J^{Po})| + S_0 = |C_W(J^{Po})| + k = |C(J_{k,r}^{Po})|$. Thus, the probability $|T| = r$ is given by:

$$\frac{1}{|G|} \left(\sum_{v \in V_L} P(|C_v(J^{Po})| = j) + \sum_{k \geq 1, r \geq 0} k Q_{k,r} P(|C(J_{k,r}^{Po})| = j) \right) = EN_j(J^{Po}) \quad (6.40)$$

$\square$

Hence, we have reduced the study of $N_j(J^{Po})$ to studying the distribution of $|T|$, the exploration process, provided that $N_j(J^{Po})$ concentrates around its mean. This is shown via a bound on its variance, followed by a Chebyshev's inequality argument. We shall postpone the proofs, and assume for now that N_k concentrates around its expectation, which we showed had the same distribution as the random walk T . To reiterate, T is implicitly defined on the Poissonized random graph $J^{Po}(\mathcal{G})$, rather than the original $J(\mathcal{G})$, so a correction will be necessary.

6.3.2 The Branching Process: from T to $\mathcal{S}$

As is classically done in the branching process approximation of the exploration process random walk, one can argue that the exploration process initially behaves like a tree: with high probability, the newly discovered vertices do not overlap with the previously explored vertices. The following two lemmas make precise the intuition that the exploration process does not overlap with itself with high probability, and hence motivates the exact pairing with a sesqui-type process.

Motivating the descendant distribution $((Y, Z))$ and the starting distribution $((Y^0, Z^0))$

Lemma 10 (Exploration is tree-like initially with good probability). *Let $\mathcal{G}$ be a good parameter list, and $T(\mathcal{G}) = (M_j, S_j)$ be the random walk associated with the exploration process. Independently for each k, r , let $\xi_{k,r}(\mathcal{G})$ be a random multiset where the tuples $g \in (V_L)^{r-1}$ appear according to independent Poisson processes with rate $rQ_{k,r}/|V_L(\mathcal{G})|^r$ (Yes, it is $r-1$, because we are exploring from one of the r vertices of the hyperedge). Condition on $(M_i, S_i)_{0 \leq i \leq j-1}$. Then, there is a coupling of (M_j, S_j) with $\xi_{k,r}$ such that with probability one,*

$$\begin{aligned} M_j - M_{j-1} &\leq \sum_{k \geq 0, r \geq 2} \sum_{(w_1, \dots, w_{r-1}) \in \xi_{k,r}} \sum_{1 \leq h \leq r-1} |C_{w_h}(H_L(\mathcal{G}))| \\ S_j - S_{j-1} &\leq \sum_{k, r \geq 1} k |\xi_{k,r}| \end{aligned} \quad (6.41)$$

and if $M_{j-1} \geq j$ (i.e. the process has not stopped), then equality holds with at least $1 - O((\log n)^{100} M_{j-1}/|V_L|)$.

Proof. As the inequality trivially holds for $M_{j-1} < j$ (the process has stopped), let us assume $M_{j-1} \geq j$. The inequality is the easy part: each $g \in \mathcal{H}_{j,k,r}$ can be mapped into a multiset of $r-1$ -tuples by deleting the first appearance of v_j . Let the resulting set be $\hat{\mathcal{H}}_{j,k,r}$. As each hyperedge occurs at a Poisson rate with $r \cdot \mu_{k,r}$, and at most r hyperedge in $\mathcal{H}_{j,k,r}$ maps to a given hyperedge in $\xi_{k,r}$, standard superposition properties of the Poisson processes gives a natural coupling with $\hat{\mathcal{H}}_{j,k,r} \subset \xi_{k,r}$.

Then, by definition, $M_j - M_{j-1} = |\bigcup_{k,r,\bar{w} \in \hat{\mathcal{H}}_{j,k,r}} \bigcup_{1 \leq j \leq r-1} C_{w_h}(H_L) \setminus (A_{j-1} \cup E_{j-1})|$, and $S_j - S_{j-1} = \sum_{k,r \geq 1} k |\hat{\mathcal{H}}_{j,k,r}|$. Hence, as $\hat{\mathcal{H}} \subset \xi$, both inequality holds.

The tricky part of this proof is proving that equality in fact holds with high probability. Let $X = \sum_{k,r} |\xi_{k,r}|$ and $X' = \sum_{k,r} (r-1) |\xi_{k,r}|$. Let us iterate over the vertices v appearing in all of the $\xi_{k,r}$, ranging over all k, r . From the coupling above, the equality holds as long as each $C_v(H)$ is pairwise disjoint, and also disjoint from $A_{j-1} \cup E_{j-1}$. (Each X' hyperedge vertices are new and lead to disjoint components in H_L .) Let us exploit the fact that $\mathcal{G}$ is nice. Recall $\alpha = (\log n)^2$, $X' \leq (\max_{Q_{k,r} \neq 0} r) X \leq \alpha X$, and all components of H_L also have size at most α .

Now, let us first condition on the size of $|\xi_{k,r}|$: then, the exact vertices in $\xi_{k,r}$ are uniformly chosen, and we shall iterate from $1 \leq \tau \leq X'$. Then, the probability of a “bad event” is bounded above by the sum of either picking one of the $M_{j-1} = |A_{j-1} \cup E_{j-1}|$ vertices, or one of the vertices in the components selected so far, which is at most $(\tau-1)\alpha \leq X'\alpha \leq \alpha^2 X$. Hence, the probability for each τ of a bad event is at most $\frac{M_{j-1}}{|V_L|} + \frac{\alpha^2 X}{|V_L|}$, and a union bound over $1 \leq \tau \leq X'$ yields that the probability of a bad event conditional on the size of $|\xi_{k,r}|$ is at most:

$$E \left(\alpha X \cdot \left(\frac{M_{j-1}}{|V_L|} + \frac{\alpha^2 X}{|V_L|} \right) | (|\xi_{k,r}|)_{k,r} \right) \quad (6.42)$$

Taking the expectation, we have that the probability of a bad event is at most $\alpha^3 (M_{j-1} EX + EX^2)/|V_L|$. As X is Poisson, $\text{Var} X = EX$, and as $E|\xi_{k,r}| = rQ_{k,r}/|V_L| \leq \alpha^2$, so $EX \leq O(\alpha^4)$, and similarly $EX^2 = O(\alpha^8)$, and as $M_{j-1} \geq j$, we can obtain the bound, as desired. $\square$

As an easier, almost special case of the above, we can describe the starting distribution (M_0, S_0) also in terms of adding independent random variables:

Lemma 11. Let (Y_0, Z_0, R) have the following distribution on $\mathbb{N}^3$ (as always, $\mathbb{N}$ starts from 0):

$$P((Y_0, Z_0, R) = (y, z, r)) = \frac{N_y(\mathcal{G})I(y > k, (z, r) = 0) + zQ_{z,r}(\mathcal{G})I(y = 0, z \geq 1)}{|\mathcal{G}|} \quad (6.43)$$

Then, there is a coupling such that:

$$M_0 \leq Y_0 + \sum_{1 \leq h \leq R} |C_{w_h}(H_L(\mathcal{G}))| \quad (6.44)$$

$$S_0 = Z_0$$

with equality holding with probability at least $1 - \alpha^3/|V_L|$.

Hence, the above two lemmas suggest a natural sesqui-type branching process estimation, $\mathcal{S}_t$. Let N' be the distribution of component size in H_L , i.e. $C_{w_h}(H_L)$. Intuitively, the second lemma tells us we should begin our process by starting at $v \in V_L$ with one copy of N' , or start at $v \in V_S$ and add R copies of N' . The first lemma tells us that subsequent iterations of the exploration process can be coupled with adding $\xi_{k,r} \sim \text{Pois}(rQ_{k,r}/|V_L(\mathcal{G})|^r)$ copies of N' for each k, r . The evolution of S_j is described in the same way.

Now, one can replace N_k and the rate $(rQ_{k,r}/|V_L(\mathcal{G})|^r)$ by the idealized distribution $P(N = k) = I(k > K)\rho_k(t_0)$ and $\lambda_{k,r}(t) = rq_{k,r}(t)/\rho_\omega(t_0)$. These are idealized distributions independent of $\mathcal{G}$, which are similar to the original N' and $|\xi_{k,r}|$, provided $\mathcal{G}$ is nice.

Defining the idealized process $\mathcal{S}_t$

Summarizing the discussion above, let us define the following sesqui-type branching process.

Definition 11. Let N be a distribution with $P(N = k) = I(k > K)\rho_k(t_0)/\rho_\omega(t_0)$ and $\lambda_{k,r}(t) = rq_{k,r}(t)/\rho_\omega(t_0)$. Let the idealized starting density satisfy:

$$P((Y_{0,t}, Z_t^0, R_t) = (y, z, r)) = \rho_y I(y > K, z, r = 0) + zq_{z,r} I(y = 0, z \geq 1) \quad (6.45)$$

and let $Y_t^0 = Y_{0,t} + \sum_{h=1}^R N_h$, where N_h are independent copies of N .

Let the sesqui-type branching process $\mathcal{S}_t$ be defined as follows. We begin with Y_t^0, Z_t^0 fertile and infertile particles. Each fertile particle gives birth to Y_t fertile particles and Z_t infertile particles (and subsequently becomes infertile), with the distribution (Y_t, Z_t) given by :

$$(Y_t, Z_t) = \left(\sum_{k \geq 0, r \geq 2} \sum_{j \leq H_{k,r,t}} \sum_{h \leq r-1} N_{k,r,j,h}, \sum_{k,r \geq 1} k H_{k,r,t} \right) \quad (6.46)$$

with $N_{k,r,j,h}$ independent copies of N and $H_{k,r,t}$ independent copies of $\text{Pois}(\lambda_{k,r}(t))$. Let $\mathcal{S}_t^1$ be the branching process starting from only one fertile particle. Finally, denote $|\mathcal{S}_t| \subset \mathbb{N} \cup \{\infty\}$ as the total number of particles of the process.

To unpackage the notation, $H_{k,r,t}$ is an approximation for the number of (k, r) hyperedges introduced for the active vertex v . $N_{k,r,j,h}$ then estimates, for each (k, r) hyperedge j , the h -th H_L component introduced to the neighborhood. As we have discussed in the proofs of the previous lemmas, intuitively the coupling relies on all the $N_{k,r,j,h}$ components being distinct/disjoint and not intersecting with the explored/active vertices.

The lemmas proved below concerning the “tree-like” behavior of the exploration process, along with Theorem 34, will lead to the proof of the following theorem:

Theorem 35 (Approximating the expectation of N_j and $N_{\geq j}$). *There exists an absolute constant D such that*

$$\begin{aligned} |EN_j(J(\mathcal{G})) - P(|\mathcal{S}_t| = j)n| &= O(j(\log n)^D n^{1/2}) \\ |EN_{\geq j}(J(\mathcal{G})) - P(|\mathcal{S}_t| \geq j)n| &= O(j(\log n)^D n^{1/2}) \end{aligned} \quad (6.47)$$

Note that the result is no longer on the Poissonized graphs $J^{Po}(\mathcal{G})$, but the original graph $J(\mathcal{G})$. As Theorem 34 concerns J^{Po} , we need to de-Poissonize the result, which we will return to later. Before we embark on the proof, we need the following coupling result.

Lemma 12 (Coupling of the Exploration Process and the Branching Process). *Let $t \in [t_0, t_1]$, and let $\mathcal{G}$ be a t -nice parameter list. There is a coupling of T , the exploration process random walk on $J^{Po}(\mathcal{G})$ and $\mathcal{S}_t$ such that for every $\Lambda(n)$, there is a constant D such that with probability at least $1 - O(\Lambda(\log n)^D n^{-1/2})$ we have $|T| = |\mathcal{S}_t|$, or $\min\{|T|, |\mathcal{S}_t|\} > \Lambda$.*

Proof. We first need basic bounds concerning the goodness of approximation from the idealized distribution to the original distribution. Recall $\alpha = (\log n)^2$. I claim there is an absolute constant D such that:

$$\begin{aligned} d_{TV}(N, N') &= O((\log n)^D n^{-1/2}) \\ \sum_{k,r \geq 0} d_{TV}(H_{k,r,t}, |\xi_{k,r}|) &= O((\log n)^D n^{-1/2}) \\ d_{TV}((Y_{0,t}, Z_t^0, R_t), (Y_0(\mathcal{G}), Z_0(\mathcal{G}), R(\mathcal{G}))) &= O((\log n)^D n^{-1/2}) \\ \sum_{k,r \geq 0: k+r \leq \alpha} P(H_{k,r,t} \geq \alpha) &\leq n^{-\omega(1)} \\ |\xi_{k,r}| &= 0 \text{ whenever } k+r \geq \alpha \end{aligned} \tag{6.48}$$

Recall $P(N = k) = \rho_k(t_0)/\rho_\omega(t_0)$, while $P(N' = k) = N_k/|V_L|$. By definition of $\mathcal{G}$ being nice, $\sum_{k > K} \rho_k(t_0) = \rho_\omega(t_0) > 0$, and each N_k is within $(\log n)^D n^{1/2}$ of $\rho_k(t_0)n$. Finally, to sum the tails, note $N_k(\mathcal{G}) = 0$ for $k > \alpha$, while $\sum_{k > \alpha} \rho_k(t_0) = n^{-\omega(1)}$ thanks to subexponentiality. The first inequality easily follows.

Next, note that $d_{TV}(Po(x), Po(y)) \leq |x - y|$, and then the proof follows from $Q_{k,r}$ being close to $q_{k,r}$ due to $\mathcal{G}$ being nice. The proof for the third inequality is analogous.

Next, as $E[H_{k,r,t}] = r q_{k,r}/\rho_\omega(t) = O(re^{-b(k+r)})$, so the probability bound follows from Chernoff bounds. The final identity is obvious from $E|\xi_{k,r}| = 0$ for $k+r \geq \alpha$.

With the bounds in mind, we can prove the coupling of the two processes inductively from 0 to Λ , using lemma 10. Note that the probability bound is vacuous for $\Lambda = \Omega(n^{1/2})$, so assume that $1 \leq \Lambda = O(n^{1/2})$. Unsurprisingly, we shall couple the numbers of type L and S particles of $\mathcal{S}_t$ with the V_L and V_S vertices found by the exploration process, from $0 \leq j \leq \Lambda$.

For the base case, by lemma 11, with probability $1 - O(\alpha^3/n)$, we have $M_0 = Y_0(\mathcal{G}) + \sum_{1 \leq h \leq R} |C_{w_h}(H_L(\mathcal{G}))|$ and $S_0 = Z_0(\mathcal{G})$ (as $|V_L| = \Theta(n)$). Now, for each copy of $C_{w_h}(H_L(\mathcal{G}))$, which is a draw from N' , we pair it with N with $N = N'$ with probability at most d_{TV} . As there are at most α copies, we obtain that the initial coupling fails with probability at most:

$$O(\alpha^3/n) + d_{TV}((Y_{0,t}, Z_t^0, R_t), (Y_0(\mathcal{G}), Z_0(\mathcal{G}), R(\mathcal{G}))) + \alpha \cdot d_{TV}(N, N') = O((\log n)^D n^{-1/2}) \tag{6.49}$$

The rest of the coupling follows from induction. By Lemma 10, equality is satisfied with probability $1 - O(\alpha^C M_{j-1}/n) = 1 - O(\alpha^C n^{-1/2})$, and then one can similarly pair $H_{k,r,t}$ with $|\xi_{k,r}|$, unless $\max_{k+r \leq \alpha} H_{k,r,t} \geq \alpha$, which occurs with probability at most $n^{-\omega(1)}$, and then pair at most $(\alpha + 1)^2 \alpha \leq \alpha^4$ independent copies of $N' \sim |C_w(H_L)|$ with N , and union bounding gives that the coupling fails with probability at most $O((\log n)^D n^{-1/2})$, as desired.

As we are only considering $\Lambda + 1$ steps in total, the union bound gives us the result. $\square$

Proof of Theorem 35

The proof of Theorem 35 now follows easily. By Theorem 12 and 34, we obtain:

$$|EN_j(J^{Po}) - P(|\mathcal{S}_t| = j)n| = |P(|T| = j) - P(|\mathcal{S}_t| = j)| \cdot n = O(j(\log n)^D n^{1/2}) \tag{6.50}$$

The rest of the proof follows by relating $EN_j(J^{Po})$ and $EN_j(J)$ via de-Poissonization. The random graph $J(\mathcal{G})$ and $J^{Po}(\mathcal{G})$ only differ in the exact number of k, r components. Note that adding an extra (k, r) will

change N_j by at most $k + 2jr$ (k for the extra k component, and then potentially the formation of a new j -component for each of the r edges), and hence we obtain:

$$|N_j(J) - N_j(J^{Po})| \leq \sum_{k,r} (2jr + k) \cdot |Pois(Q_{k,r}(\mathcal{G})) - Q_{k,r}(\mathcal{G})| \quad (6.51)$$

The expectation of the deviation from the mean can be bounded by $\sqrt{Var(Pois)} = \sqrt{\mu}$, and as $Q_{k,r} \leq Be^{-b(k+r)}n$ by niceness, it follows $|EN_j(J) - EN_j(J^{Po})| = O(jn^{1/2})$, which gives us the theorem for EN_j . The proof for $EN_{\geq j}$ follows analogously.

Hence, we have related the expectation of the component sizes $N_k(J(\mathcal{G}))$ to the distribution of the branching process $\mathcal{S}_t$. As we have defined $\mathcal{S}_t$ independently of $\mathcal{G}$, the concentration of N_k to ρ_k then gives us the following corollary:

Corollary 11. $\rho_k(t) = P(|\mathcal{S}_t| = k)$, and $\rho(t) = P(|\mathcal{S}_t| = \infty)$.

Proof. Note that with probability $1 - o(1)$, we have that $\mathcal{G}_{tn}$ is nice, and in particular $|EN_k - \rho_k n| \leq (\log n)^2 n^{1/2}$. Furthermore, by Theorem 35, we obtain: $|EN_k(tn) - P(|\mathcal{S}_t| = k)n| = O(k(\log n)^D n^{1/2})$. Putting the two together gives us:

$$|P(|\mathcal{S}_t| = k) - \rho_k(t)| = O(k(\log n)^D n^{-1/2} + (\log n)^2 n^{-1/2}) = o(1) \quad (6.52)$$

As $\mathcal{S}_t$ and ρ_k are defined independently of n , we obtain equality.

As for the result on $\rho(t) = P(|\mathcal{S}_t| = \infty)$, [26] gives a proof that $\rho(t) = 1 - \sum_{k \geq 1} \rho_k(t)$ for Achlioptas processes. Hence, the result follows naturally from $\rho_k = P(|\mathcal{S}_t| = k)$. $\square$

6.3.3 Analysis of the Branching Process $\mathcal{S}$

Hence, we have reduced our analysis of the component size distribution N_j to the study of the sesqui-type branching process family $\mathcal{S}_t$. In Appendix E, we give a sketch of the proof that $\mathcal{S}_t$ is indeed a t_c -critical branching family. Intuitively, what that means is that the probability generating function is analytic in t, α, β , with $E[Y_{t_c}] = 1$, and $E[Y]$ locally increasing. The key difficulty in the proof is analytical, and relies on the use of partial differential equations and the explicit differential equations derived for $q_{k,r}$.

Then, our results of Chapter 4, Lemmas 4 and 5 can be summarized as follows:

Theorem 36. $\mathcal{S}_t$ the branching process coupled with the random graph exploration process, satisfies the following:

1. there exist constants $\epsilon_0, c > 0$ and analytic functions θ and ψ on $[t_c - \epsilon_0, t_c + \epsilon_0]$ such that:

$$P(|\mathcal{S}| = k) = (1 + O(1/k))k^{-3/2}\theta(t)e^{-k\psi(t)} \quad (6.53)$$

with $\theta > 0$, $\psi \geq 0$, with $\psi(t_c) = \psi'(t_c) = 0$ and $\psi''(t_c) > 0$.

2. The survival probability $\rho(t) = P(|\mathcal{S}_t| = \infty) = 0$ for $t_c - \epsilon \leq t \leq t_c$, and is positive for $t_c < t \leq t_c + \epsilon_0$.
3. $\rho(t)$ is analytic on $[t_c, t_c + \epsilon_0]$.

From the point probabilities, one can obtain a useful asymptotic formula for the tail probabilities of $|\mathcal{S}_t|$.

Corollary 12. There exists constants ϵ_0, c_0 , and C such that for $t = t_c + \epsilon$, with $\epsilon > 0$ and $\epsilon < \epsilon_0$,

$$P(|\mathcal{S}_t| \geq k) - P(|\mathcal{S}_t| = \infty) \leq C\epsilon^2 k^{-3/2} \exp(-kc_0\epsilon^2) \quad (6.54)$$

Proof. In the neighborhood of t_c , i.e. $t_c + \epsilon$, $|\epsilon| < \epsilon_0$, there is a $c_0 > 0$ such that $\psi(t_c + \epsilon) \geq c_0\epsilon^2$. Then, by the subexponentiality of the point probabilities,

$$\sum_{t=k}^{\infty} P(|\mathcal{S}| = t) \leq (1 + O(1/k))k^{-3/2}\theta(t)e^{-k\psi(t)} \frac{1}{1 - e^{-\psi(t)}} \leq 2\Theta c_0\epsilon^2 k^{-3/2} e^{-kc_0\epsilon^2} \quad (6.55)$$

as desired, where Θ is the minimum of $\theta(t)$ on $t_c \pm \epsilon_0$. $\square$

By setting $k = \epsilon^{-2} \cdot \omega(1)$, with $\omega(1)$ growing slowly to ∞ , the above corollary implies that $P(|\mathcal{S}_t| \geq \omega(1)\epsilon^{-2}) = P(|\mathcal{S}_t| = \infty) + o(1)$. This is highly reminiscent of our previous results for Erdős-Rényi graphs, where eventually there does not exist other components of order $> \epsilon^{-2}$ other than the giant component.

6.4 Putting Things Together: The Final Proof

We are finally ready to prove the main results. First, we shall prove Theorem 26.

6.4.1 Small Components

First, we shall apply McDiarmid's inequality to prove that $N_k(J)$ is concentrated around its expectation:

Lemma 13. *Given $\mathcal{G}$ a t -nice parameter list, with probability at least $1 - n^{-\omega(1)}$ we have $|N_k(J) - EN_k(J)| \leq k(\log n)n^{1/2}$ and similarly for $N_{\geq k}$ for all of $1 \leq k \leq n$.*

Proof. The proof is a simple application of McDiarmid's inequality. The random variables are the choices of each of the stubs of the (k, r) hyperedges, of which there are $\sum_{k,r} rQ_{k,r} \leq \sum_{k,r} rBe^{-b(k+r)}n = O(n)$.

The function is of course the random quantity N_k , and the bounded-differences condition is easily satisfied, as a removal or addition of a single stub results in at most $2k$ change in N_k . Hence, the change resulting from a single random variable is at most $4k$, and hence McDiarmid's inequality yields:

$$P(|N_k(J) - EN_k(J)| \geq k(\log n)n^{1/2}) \leq 2 \exp\left(-\frac{2(k(\log n)n^{1/2})^2}{O(n)(4k)^2}\right) \leq n^{-\omega(1)} \quad (6.56)$$

A similar inequality holds for $N_{\geq k}$ from $\mathcal{G}$ with an almost identical argument. $\square$

A simple union bound yields the following corollary:

Corollary 13. *With probability at least $1 - O(n^{-99})$, the following hold for all $i_0 \leq i \leq i_1$ and all $k \geq 1$:*

$$\begin{aligned} |N_k(i) - P(|\mathcal{S}|_{i/n} = k)n| &\leq k(\log n)^D n^{1/2} \\ |N_{\geq k}(i) - P(|\mathcal{S}|_{i/n} \geq k)n| &\leq k(\log n)^D n^{1/2} \end{aligned} \quad (6.57)$$

Proof. Suffices to assume $k < n$. By applying Theorem 35 with Lemma 13, the triangle inequality yields:

$$\begin{aligned} |N_k(i) - P(|\mathcal{S}|_{i/n} = k)n| &\leq k(\log n)^D n^{1/2} \\ |N_{\geq k}(i) - P(|\mathcal{S}|_{i/n} \geq k)n| &\leq k(\log n)^D n^{1/2} \end{aligned} \quad (6.58)$$

hold simultaneously for $k = 1$ to n with probability $1 - n^{-\omega(1)}$. Union-bounding over $O(n)$ instances of $i_0 \leq i \leq i_1$ gives the corollary. $\square$

Now, to prove the local asymptotics result, first, Theorem 36 indeed gives us:

$$\rho_k(t) = (1 + O(1/k))k^{-3/2}\theta(t)e^{-\psi(t)k} \quad (6.59)$$

To prove the uniform convergence of N_k to ρ_k , note that by Theorem 36, $a = \psi''(t_c) > 0$, $\psi(t_c) = \psi'(t_c) = 0$, and $b = \theta(t_c) > 0$, so by shrinking ϵ_0 if necessary, one can bound $\psi \leq a\epsilon^2$, $\theta(t) \geq b/2$, and obtain, for $1 \leq k \leq n^{1/10}$ and $a\epsilon^2 k \leq 0.1 \log n$, we have: $\frac{k n^{1/2}}{k^{-5/2} e^{-\psi(t)k} n} = o(n^{-1/30 - 0.00001})$. Hence, the $k \log n n^{1/2}$ error term will be dominated by $\rho_k n$, and in particular $k(\log n)^D n^{1/2} - o(1/k)\rho_k(t)n$. Then, Corollary 13 yields the result, as desired. The proof for $N_{\geq k}$ proceeds similarly.

6.4.2 Size of the largest component

Next, we shall prove Theorem 28, the weakened version of the global convergence theorem. Instead of proving uniform convergence to $\rho(t)$, we shall show that $L_1(i) = o(n)$ at subcritical t , while $L_1(tn)/n \mapsto \rho(t)$, which is analytic and increasing for supercritical t . In both the subcritical case and the supercritical case, the first step is to choose a Λ such that $N_{\geq \Lambda} = P(|S_{i/n}| \geq \Lambda)n + \Lambda(\log n)^D n^{1/2}$.

For the subcritical case, note that trivially $L_1 > \Lambda$ implies $L_1 \leq N_{\geq \Lambda}$. Hence, if we can find Λ such that $N_{\geq \Lambda}$ is small, we obtain $L_1 < \max\{\Lambda, N_{\geq \Lambda}\}$. Recall that one can bound via Theorem 13 that with high probability:

$$N_{\geq \Lambda} \leq \Lambda(\log n)^D n^{1/2} + nP(|S_{i/n}| \geq \Lambda) \quad (6.60)$$

Hence, by letting $\Lambda \gg \epsilon^{-2} \log n$, one obtains that $P(|S_{i/n}| \geq \Lambda) \leq P(|S_{i/n}| = \infty) + n^{-\omega(1)} = n^{-\omega(1)}$, as the process is subcritical. Hence, one obtains that with high probability, for $\Lambda \gg \epsilon^{-2} \log n$, which is $O(\epsilon^{-2}(\log n)^{D+1}n^{1/2})$, as desired.

6.4.3 Size of the largest component: Supercritical Case

For the supercritical case, we need a slight modification of the Erdős-Rényi sprinkling technique.

Theorem 37 (Sprinkling, Achlioptas version). *There exists absolute constants λ and η such that conditional on $N_{\geq \Lambda}(i) \geq x$, we have $L_1(i + \lambda n^2/(\xi \Lambda x)) \geq (1 - \xi)N_{\geq \Lambda}(i)$ with probability at least $1 - \exp(-\eta x/\Lambda) - n^{-\omega(1)}$.*

Proof. We shall assume $\xi \in (0, 1)$. As $N_{\omega}(i)$ is monotone in i , and N_{ω} is $\Theta(n)$ for $t > 0$, there is a constant C such that with probability $1 - n^{-\omega(1)}$,

$$\min_{i \geq i_0} N_{\omega}(i) \geq N_{\omega}(i_0) \geq Cn \quad (6.61)$$

Now, let W be the union of all components of G_i with size at least Λ . Thus, the number of components of G_j meeting W is at most $|W|/\Lambda$ and decreases monotonically. Also, until there is a component containing at least $(1 - \xi)|W|$ vertices, we have at each step probability at least:

$$\min_{j \geq i} \left(\frac{N_{\omega}(j)}{n} \right)^{\ell-2} \frac{|W|}{n} \frac{\xi|W|}{n} \geq C^{\ell-2} \xi \left(\frac{|W|}{n} \right)^2 = q \quad (6.62)$$

of joining two disjoint components meeting W . (This is because if we select all components of size ω , the bounded-size Achlioptas rule is blind to which components to add).

Now, let $M = \frac{4n^2}{\alpha^{\ell-2}\xi\Lambda x}$, and note that $qM/2 = 2|W|^2/(\Lambda x)$, and as $|W| = N_{\geq \Lambda}(i) \geq x \geq \Lambda$, we have $qM/2 = 2|W|^2/(\Lambda x) > \max\{2x/\Lambda, |W|/\Lambda\}$. For the sake of completeness, we shall state the standard Chernoff bound:

Lemma 14. *Let X_i , $1 \leq i \leq M$, be a sequence of Bernoulli random variables. Let $P(X_j | X_{1:j-1}) > q$. Then, there is an absolute constant $\eta > 0$ such that: $P(\sum_{i=1}^M X_i < qM/2) < \exp(-\eta qM)$*

We defer the proof to Appendix D. By Chernoff, after M steps, with probability at least $1 - \exp(-\eta qM) = 1 - \exp(-\eta qM) \geq 1 - \exp(-\eta' x/\Lambda)$, we have a component containing at least $(1 - \xi)|W|$ vertices, or we have joined at least $qM/2$ distinct components intersecting W . As there are at most $|W|/\Lambda$ disjoint components intersecting W , the latter is impossible, as desired. $\square$

Consequences of Sprinkling

Let $i^* = t_c n + (1 - \xi)\epsilon n$, with $\xi = o(1)$. Then, Theorem 13 then yields:

$$P(|N_{\geq \Lambda}(i^*) - P(S_{i^*/n} \geq \Lambda)n| \leq \Lambda(\log n)^D n^{1/2}) \quad (6.63)$$

Now, let us set $\epsilon n \gg \Lambda \gg (\xi\epsilon)^{-2}$. As we have assumed that $\epsilon^3 n \mapsto \infty$, we can eventually achieve this. By lemma 12, $P(\mathcal{S}_{i^*/n} \geq \Lambda) = P(\mathcal{S}_{i^*/n} = \infty) + O(e^{-c_0\xi^{-2}})$. Putting the identities together, one obtains that with high probability,

$$N_{\geq \Lambda}(i^*) = P(\mathcal{S}_{i^*/n} = \infty)n \pm O(\Lambda(\log n)^D n^{1/2} + e^{-c_0\omega}n) \quad (6.64)$$

Note that as $\mathcal{S}$ is supercritical at $t = t_c + \epsilon$, $P(|\mathcal{S}_{i^*/n}| = \infty) = \Theta(\epsilon n)$, and by the argument above, $EN_{\geq \Lambda} = \Theta(\epsilon n)$ as well.

To prove the concentration of $L_1(i)$, note that the upper bound for $L_1(i)$ is immediate from $L_1(i) \leq N_{\geq \Lambda}(i)$, which, as we just demonstrated, converges (after normalization) to $P(|\mathcal{S}_t| = \infty)$.

The lower bound, on the other hand, will follow from the sprinkling argument. As we have just shown that $N_{\geq \Lambda}(i^*) \geq x = \Theta(\epsilon n)$, we note that $\lambda n^2/(\xi\Lambda x) = o(\xi\epsilon n)$, $x/\Lambda = \Theta(\epsilon n/\Lambda) = \omega(1)$, and $i^* + \frac{\lambda n^2}{\xi\Lambda x} \leq i$, so by the sprinkling lemma, with high probability,

$$L_1(i) \geq L_1(i^* + \frac{\lambda n^2}{\xi\Lambda x}) \geq (1 - \xi)N_{\geq \Lambda}(i^*) \geq (1 - \xi) \left(P(|\mathcal{S}_{t_c+\epsilon}| = \infty)n + o(n) - O(\Lambda(\log n)^D n^{1/2} + e^{-c_0\omega}n) \right) \quad (6.65)$$

where the last inequality follows from the continuity of $P(|\mathcal{S}_t| = \infty)$ with respect to t , and the fact that $(i - i^*) = o(n)$. As ξ is arbitrary and $o(1)$, as $\xi \mapsto 0$, we thus obtain our concentration result for $L_1(i) \mapsto \rho(i/n)$. Finally, the other properties regarding ρ , such as analyticity, the positivity of the right derivative at criticality, etc, derives from the survival probability results of Theorem 36.

6.5 Summary and Takeaway

In this chapter, we have proven that bounded-size Achlioptas processes form a type of universality class for Erdős-Rényi phase transition behavior. In other words, as long as the candidate edges are selected uniformly at random, and the edge selection rule is indifferent to component sizes beyond a certain threshold K , a random graph process with fairly general edge selection rule exhibits phase transition behavior similar to that of the Erdős-Rényi process we have studied in Chapter 2. Of course, the precise constants, such as the criticality time t_c and the slope/rate of increase of the giant component beyond criticality, depend on the precise Achlioptas decision rule.

Putting aside the technical aspects of the proof, such as the myriad inequalities and ϵ -bounding arguments, the basic structure of the proof relies on the following three components:

- Differential Equations governing normalized component distribution
- The Neighborhood Exploration Process
- The Idealized Branching Process

The first component, which are the idealized quantities ρ_k and $q_{k,r}$, follow well-behaved differential equations implied by the Achlioptas rule R . Typically, the differential equations are polynomial and hence smooth, with the derivative of each function depending only on its predecessors. The study of these idealized functions ρ_k and $\rho = 1 - \sum_{k \geq 1} \rho_k$ provide a “macroscopic” view of the random graph process.

On the other hand, the neighborhood exploration process for the random graph G provides a “microscopic” view of the random graph process. What is needed as an ingredient for the exploration process is the component size distribution, which is often given by the solution to the aforementioned differential equations. In our proof, the graph parameter list $\mathcal{G}$, and its concentration to ρ_k and q , played such a role.

While the neighborhood exploration process gives a natural way of analyzing the microscopic evolution of component size distribution for “small” time windows, one can analyze this process by coupling it with a natural branching process. Typically, this branching process is motivated by assuming that initially, the neighbors introduced by the exploration process do not overlap with the previously explored vertices.

The branching process associated to the neighborhood exploration process gives a striking and intuitive picture of the phase transition in random graph processes. It is well known that a branching process becomes critical when the expected number of offspring becomes 1. Thus, the random graph process becomes critical when the expected number of new neighbors reaches 1. Intuitively, as the random graph grows, there exist many more vertices in increasingly larger components. If potential edges are selected randomly, the expected number of new neighbors will continue to rise until criticality.

As the main focus of our paper was on phase transition behavior of the largest component size L_1 , we have focused our analysis on ρ , which is given by the survival probability of the associated branching process. However, we have also proven the uniform convergence of local statistics N_k , and properties of the normalized family of distributions ρ_k . The precise asymptotics of ρ_k was instrumental in proving results concerning $\rho = 1 - \sum_{k=1}^{\infty} \rho_k$, and it remains to be seen if the local information of the component size distribution can yield other interesting insights into other graph properties, such as component complexity (are the small components trees, unicyclic, or (k, d) components for $d > 0$?).

Finally, the general method of employing differential equations, neighborhood exploration processes, and coupled branching processes can easily extend to the analysis of other general random graph processes. [13] gives an example of an analysis of the CHKNS model (by Callaway, Hopcroft, Kleinberg, Newman, and Strogatz) using $N_k(t)$, which is the expected number of components of size k , which by a factor of k is related to our N_k in the analysis of Achlioptas processes. The DEM can be used to obtain the differential equations for N_k , which can then be used to prove the phase transition and find the critical value for the process. Hence, the proof given in this chapter merits an independent methodological importance by introducing powerful techniques that can be used to analyze more general graph processes.

6.6 Conclusion and Future Work

Starting from Bollobás' seminal results concerning the giant component phase transition for Erdős-Rényi processes, we have built up the machinery needed to extend the phase transition result to general bounded-size Achlioptas processes. While we have not obtained the detailed and granular picture of the random graph process as we have for Erdős-Rényi processes (such as component complexity, trees, etc), we nevertheless were able to generalize many of the salient features of the Erdős-Rényi phase transition to the Achlioptas process setting. In doing so, we have highlighted the techniques of branching process methods as well as differential equation methods, which have wide applicability in analyzing randomized discrete processes.

It is important to emphasize that we have sacrificed a lot of the strength present in the original work by Riordan and Warnke to obtain a simpler exposition. In particular, they not only proved pointwise convergence of $L_1(G_{tn})/n$ to $\rho(t)$, but uniform convergence in the interval $i \in [i_0, i_1] = (t_c \pm \sigma)n$, up to a multiplicative factor $(1 \pm \log(\epsilon^3 n)^{-1/2})$, which is closer to the spirit of the DEM-type results.

Furthermore, the detail that is missing in this exposition is precisely the second-moment bounding of $N_{\geq \Delta}$, the number of vertices in "large components". This is similar in spirit to Theorem 3, which uses variance bounds and Chebyshev's inequality to give tighter bounds (Of course, the precise argument is much more complicated and involves sandwiching the original graph between two Poissonized graphs). Hence, one can draw further analogies of Riordan and Warnke's proof to Bollobás' proof.

In light of the universality result for giant component phase transition, one can ask the natural question: what about unbounded-size rules? Achlioptas originally considered the product-rule, the rule that chooses the edge that minimizes the product of the component sizes connected by the two edges. While [26] proves that all Achlioptas processes have continuous phase transitions, [25] speculates that the emergence of the giant component for product-rule Achlioptas processes is faster than linear near criticality, although the precise dynamics are yet unknown.

Regardless, as demonstrated by [26], many of the features of Erdős-Rényi phase transition still hold for general Achlioptas processes, including the continuity of the normalized $L_1(G_{tn})/n$ beyond criticality. Perhaps by relaxing the host of qualitative features characterizing Erdős-Rényi phase transitions, one can obtain larger universality classes within the space of random graph processes, possibly containing other processes such as preferential attachment models, CHKNS models, and inhomogeneous random graph processes.

Bibliography

- [1] D. Achlioptas, R. M. D'Souza, and J. Spencer. Explosive percolation in random networks. *Science (New York, N.Y.)*, 323(5920), March 2009.
- [2] E. M. Airolidi, T. B. Costa, and S. H. Chan. Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. *ArXiv e-prints*, Nov. 2013.
- [3] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, Jan 2002.
- [4] D. Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *The Annals of Probability*, 25(2):812–854, 1997.
- [5] Y. Azar, A. Z. Broder, A. R. Karlin, and E. Upfal. Balanced allocations. *SIAM J. Comput.*, 29(1):180–200, Sept. 1999.
- [6] T. Bohman and A. Frieze. Avoiding a giant component. 19:75 – 85, 08 2001.
- [7] B. Bollobas. The evolution of random graphs. *Transactions of the American Mathematical Society*, 1984.
- [8] B. Bollobas. *Modern Graph Theory*. Springer, 1998.
- [9] B. Bollobas. *Random Graphs*. Cambridge University Press, 2001.
- [10] B. Bollobas, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *ArXiv Mathematics e-prints*, Apr. 2005.
- [11] A. Clauset, C. Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *ArXiv e-prints*, June 2007.
- [12] J. Diaz and D. Mitsche. The cook-book approach to the differential equation method. *Computer Science Review*, 4(3):129 – 151, 2010.
- [13] R. Durrett. *Random Graph Dynamics*. Cambridge University Press, 2007.
- [14] P. Erdős and A. Rényi. On the evolution of random graphs. In *PUBLICATION OF THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES*, pages 17–61, 1960.
- [15] W. Feller. *An Introduction to Probability Theory and its Applications Vol. I*. Wiley, 1968.
- [16] S. Janson, O. Riordan, and L. Warnke. Sesqui-type branching processes. *ArXiv e-prints*, June 2017.
- [17] M. Krivelevich and B. Sudakov. The phase transition in random graphs - a simple proof. *ArXiv e-prints*, Jan. 2012.
- [18] L. Lovász and B. Szegedy. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933 – 957, 2006.

- [19] Z. P. Mihyun Kang. Random graphs: Theory and applications from nature to society to the brain.
- [20] M. Mitzenmacher, A. W. Richa, and R. Sitaraman. The power of two random choices: A survey of techniques and results. In *in Handbook of Randomized Computing*, pages 255–312. Kluwer, 2000.
- [21] H. L. Montgomery. The pair correlation of zeros of the zeta function. *Analytical number theory*, 1973.
- [22] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64:026118, Jul 2001.
- [23] A. E. R. Peter D. Hoff and M. S. Handcock. Latent space approaches to social network analysis. *JASA, Theory and Methods*, 97, 2002.
- [24] A. Rinaldo, S. E. Fienberg, and Y. Zhou. On the geometry of discrete exponential families with application to exponential random graph models. *Electron. J. Statist.*, 3:446–484, 2009.
- [25] O. Riordan and L. Warnke. The phase transition in bounded-size achlioptas processes.
- [26] O. Riordan and L. Warnke. Explosive percolation is continuous. *Science*, 333(6040):322–324, 2011.
- [27] O. Riordan and L. Warnke. The evolution of subcritical achlioptas processes. *Random Structures and Algorithms*, 47(1):174–203, 2015.
- [28] O. RIORDAN and L. WARNKE. Convergence of achlioptas processes via differential equations with unique solutions. *Combinatorics, Probability and Computing*, 25(1):154–171, 2016.
- [29] J. Spencer and N. Wormald. Birth control for giants. *Combinatorica*, 27(5):587–628, Sep 2007.
- [30] F. Spitzer. *A Combinatorial Lemma and its Application to Probability Theory*, pages 3–19. Birkhäuser Boston, Boston, MA, 1991.
- [31] K. F. Stanley Wasserman. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.
- [32] T. Tao, V. Vu, and M. Krishnapur. Random matrices: Universality of esds and the circular law. *Ann. Probab.*, 38(5):2023–2065, 09 2010.
- [33] M. Tenenbaum and H. Pollard. *Ordinary Differential Equations: An Elementary Textbook for Students of Mathematics, Engineering, and the Sciences*. Harper’s mathematics series. Harper & Row, 1963.
- [34] G. L. R. Tom A. B. Snijders, Philippa E. Pattison and M. S. Handcock. New specifications for exponential random graph models. *Social Methodology*, 36(1), 2006.
- [35] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12, 2012.
- [36] N. C. Wormald. Differential equations for random processes and random graphs. *The Annals of Applied Probability*, 5(4):1217–1235, 1995.
- [37] N. C. Wormald. The differential equation method for random graph processes and greedy algorithms. In *Lectures on Approximation and Randomized Algorithms*, pages 73–155. PWN, 1999.
- [38] T. Łuczak. Component behavior near the critical point of the random graph process. *Random Structures and Algorithms*, 1(3):287–310, 1990.

Appendix A

Proof of Theorem 1 and other Combinatorial Results

We shall begin with a proof of a generalized version of Cayley's Theorem.

Theorem 38 (Extended Cayley's Theorem). *The number of trees with n distinct labelled vertices is n^{n-2} . More generally, denote $F(n, k)$ as the number of forests with n distinct vertices with k components, such that $1, 2, \dots, k$ belong to distinct components. Then, $F(n, k) = kn^{n-1-k}$.*

Proof. First, note that there is a bijection between the forests mentioned above and a tree obtained by appending to the forest a vertex 0, which is then connected to only 1 through k . It is easy to see that a general tree with 0 connected to only 1 through k corresponds to the appropriate forest by deleting 0.

Next, to count such trees, we use Prüfer encoding. Given a tree, remove the leaf (a vertex of degree 1) with the largest entry, and record the number of its unique neighbor, until we are left with the lone vertex 0. Given the structure of the tree, this implies that we will obtain a code of length n , with the last k entries being 0, and the $n - k$ th entry in $\{1, \dots, k\}$.

Conversely, given an arbitrary code P of length n with the last $k + 1$ entries as described above, one can construct the tree as follows:

1. Pick the largest number not in the sequence, x , and attach it to the vertex corresponding to the first entry of P .
2. Delete the first entry of P , append x to the end of P .
3. Continue the said process until n edges are added.

It is not difficult to see that the above process is the inverse of the encoding, where we add back the edges in the order that they were removed. Thus, F is bounded above by the number of possible Prüfer codes, which is easily kn^{n-1-k} . To prove equality, one must show any appropriate Prüfer code decodes into a tree with 0 being connected to 1 through k .

First, I claim that the resulting graph is a tree. One can prove this via induction on the length of the Prüfer code. The base case is trivial. Now, after the first iteration of the decoding algorithm, note that the added vertex j does not appear further in the algorithm. Hence, the remaining steps of decoding corresponds to a decoding of a tree with vertices $\{0, 1, \dots, n\} - j$. Hence, by inductive hypothesis, the subsequent steps after the first step produce a tree, and as we are adding a vertex of degree 1 to a tree, the original graph is a tree as well, as desired.

Furthermore, by construction, the degree of a vertex v is precisely 1 less than the number of appearances of v in the Prüfer code. Hence, the code P of length n with only the last k entries being 0, the $k + 1$ -th entry from the end being 1 through k , and the remaining entries being 1 through n correspond to a graph with 0 connected to 1 through k , as desired. $\square$

We are now ready to prove the theorem for $d \leq k$:

Theorem 39. 1. The number of unicyclic components $C(k, 0) = \frac{1}{2}k^{k-1} \sum_{r=3}^k \prod_{j=1}^{r-1} (1 - \frac{j}{k}) = \frac{1}{2}(k - 1)! \sum_{j=0}^{k-3} \frac{k^j}{j!} \sim (\pi/8)^{1/2} k^{k-1/2}$

2. There is a constant c such that for $1 \leq d \leq k$, $C(k, d) \leq (c/d)^{d/2} k^{k+(3d-1)/2}$

Proof. To prove the first statement, note that a unicyclic graph G can be decomposed into a unique cycle C , with r edges, and a forest F with each r vertices of C belonging to different components.

More generally, a (k, d) component, i.e. a connected graph of k vertices and $k+d$ edges, can be decomposed into a component H , whose vertices have degree at least 2, and a forest F , where each component of F contains exactly one vertex of H .

Using this basic fact, which can be proved via an induction on d (for $d \geq 0$, find a cycle and then use the inductive hypothesis on remaining components), we can prove both results.

First, for unicyclic graphs, suppose the unique cycle has r vertices. Then, there are $\binom{k}{r} \frac{r!}{2r} = \binom{k}{r} / (2r)$ choices for the cycle, and $r k^{r-1-k}$ choices for the forest, by the generalized Cayley's theorem.

Hence, we obtain:

$$C(k, 0) = \sum_{r=3}^k \frac{\binom{k}{r}}{2r} r k^{r-1-k} = \frac{1}{2} k^{k-1} \sum_{r=3}^k \prod_{j=1}^{r-1} (1 - \frac{j}{k}) \quad (\text{A.1})$$

To show asymptotics, as $\log(1-x) \leq -x$, $\prod_{j=1}^{r-1} (1 - \frac{j}{k}) \leq \exp(-\binom{r}{2}/k)$, and for $r = O(k^{2/3})$, $\prod_{j=1}^{r-1} (1 - \frac{j}{k}) = e^{-r^2/(2k)} (1 + O(r/k) + O(r^3/k^2))$.

Hence,

$$\sum_{r=3}^k \prod_{j=1}^{r-1} (1 - \frac{j}{k}) = \sum_{r=3}^{n^{1/2+\alpha}} e^{-r^2/(2n)} + O(1) = \sqrt{n} \int_0^\infty e^{-x^2/2} dx + O(1) = (\frac{\pi k}{2})^{1/2} + O(1) \quad (\text{A.2})$$

Thus, multiplying by $\frac{1}{2} k^{k-1}$ gives us the asymptotics, as desired.

As for the second result, one similarly attains the subgraph H instead of the cycle C , and the residual forest F . Let r be the number of vertices of H . Then, as there are $k-r$ edges in F (r components with k vertices), $e(H) = r + d$, and there are $r k^{k-1-r}$ possible G for each choice of H .

Counting the number of H with r vertices is difficult compared to the number of cycles. Instead, we can bound the number of H in the following way. Let t be the number of vertices of degree exactly 2 in H . Then, from each graph H , one can remove the vertices of degree 2 while maintaining the “edge” running through the vertices. Every vertex of the resulting multi-graph M (loops and multiple edges between a vertex pair can occur), has degree at least 3. From the construction, it is clear that H can be obtained from subdividing some of the edges of M .

If M has t vertices, one obtains H by adding $r - t = u$ subdivisions into each of the $t + d$ edges or loops of M . For a given M , one can count the possible subdivisions as $u! \binom{u+t+d-1}{t+d-1}$ (the “balls-and-bins” argument). This, however, is an upper-bound, as the presence of multi-edges will imply some graphs are counted multiple times.

Let $\psi(t, d)$ be the number of the aforementioned multigraphs with t vertices and $t + d$ edges. $\psi(t, d)$ can be crudely bounded by:

$$\psi(t, t + d) \leq \binom{\binom{t}{2} + t + (t + d - 1)}{t + d - 1} \quad (\text{A.3})$$

where we partition the $t + d$ edges and loops into $\binom{t}{2} + t$ possible classes for the edges and possible self-loops. As each vertex of M has degree at least 3, $3t \leq 2(t + d) \implies t \leq 2d$. Hence, the above binomial expression can be bounded above by $(c_1 d)^{t+d-1}$ via Stirling's approximation.

The preceding argument can be summarized as follows:

$$G \xrightarrow{\text{non-forest component: } rk^{k-1-r}} H \xrightarrow{\text{reduction: } \leq u! \binom{u+t+d-1}{t+d-1}} M \leq \binom{\binom{t}{2} + t + (t+d-1)}{t+d-1} \leq (c_1 d)^{t+d-1}$$

Hence, we obtain:

$$C(k, d) \leq \sum_{t=1}^{2d} \sum_{u=0}^{k-t} (t+u) k^{k-1-t-u} \binom{k-t}{u} u! \binom{k}{t} (c_1 d)^{t+d-1} \quad (\text{A.4})$$

One can further bound, in a technique reminiscent of the first part of the theorem, $\binom{k}{t} \binom{k-t}{u} u! = \frac{1}{t!} (k)_{t+u} \leq \frac{k^{t+u}}{t!} \exp(-u^2/(2k))$. Plugging this in, we obtain:

$$C(k, d) \leq k^{k-1} \sum_{t=1}^{2d} \sum_{u=0}^{k-t} \frac{1}{t!} (c_1 d)^{t+d-1} \exp(-u^2/2k) \binom{u+t+d-1}{t+d-1} (t+u) \quad (\text{A.5})$$

One can separately bound the double sum for $u \leq d$ and $d < u \leq k+t$. For small u , the binomial term is bounded crudely above by $2^{t+2d-1} \leq 4^{t+d-1}$, and $t+u \leq 3d$. Hence, up to a factor of k^{k-1} , those terms are bounded above by:

$$\sum_{t=1}^{2d} \frac{1}{t!} (c_2 d)^{t+d-1} 3d \sum_{u=0}^d \exp(-u^2/2k) \leq \sqrt{k} \sum_{t=1}^{2d} \frac{1}{t!} (c_2 d)^{t+d-1} = \sqrt{k} d^{d-1} \sum_{t=1}^{2d} \frac{(c_2 d)^t}{t!} \quad (\text{A.6})$$

where we use the integral approximation as we had in part 1.

Next, we can adjust c_2 such that the above series is strictly increasing, and in fact more than doubles at each step ($c_2 > 4$). Hence, the series is, up to a constant, bounded above by the last term, which by Stirling's approximation is $O(\frac{(c_2 d)^{2d}}{d^{2d} d^{1/2}}) = O(c_3^{2d} d^{-1/2})$. Plugging this bound back gives the upper-bound of $\sqrt{k} (c_4 d)^{d-3/2} \leq (c_5/d)^{d/2} k^{3d/2-1}$ as $d \leq k$.

As for large u , the binomial term $\binom{u+t+d-1}{t+d-1}$ is no longer negligible, and we bound this above by $\frac{u^{t+d}}{(t+d-1)!}$ (up to some exponential of d), as $u > d$.

Then, our sum is bounded above by:

$$\frac{\sum_{t=1}^{2d} (c_6 d)^{t+d-1} \sqrt{k} \sum_{u=0}^{\infty} u^{t+d} \exp(-u^2/2k)}{t!(t+d-1)!} \leq \sum_{t=1}^{2d} \frac{(c_6 d)^{t+d-1} k^{(t+d+1)/2}}{t!(t+d-1)!} \int_0^{\infty} x^{t+d} \exp(-x^2/2) dx \quad (\text{A.7})$$

The integral above can be bounded above by $((t+d)!)^{1/2}$ via a Gamma integral, and hence we obtain an upper bound of:

$$k^{(d+1)/2} \sum_{t=1}^{2d} \frac{((t+d)!)^{1/2} (c_6 d)^{t+d-1} k^{t/2}}{t!(t+d-1)!} < k^{(d+1)/2} \sum_{t=1}^{2d} \frac{(c_7 d)^{t+d} k^{t/2}}{t!((t+d)!)^{1/2}} < k^{(d+1)/2} \sum_{t=1}^{2d} \frac{(c_8 d)^{(t+d)/2} k^{t/2}}{t!} \quad (\text{A.8})$$

where the last inequality comes from Stirling's approximation. Lastly, one can increase c_8 such that the above series more than doubles at each iteration, which implies that the sum is bounded, up to a constant, by the last term, which is $k^{(d+1)/2} \frac{(c_9 d)^{3d/2} k^d}{(2d)!} \leq k^{(d+1)/2+d} (c_{10}/d)^{d/2}$.

Putting the two terms together and adding back the k^{k-1} factor, our sum is bounded by $k^{k-1} (c/d)^{d/2} k^{3d/2-1}$, as desired. $\square$

To go beyond $d \leq k$, we only need the crude bound $C(k, d) \leq \binom{k}{k+d}$ (all graphs of $k+d$ edges), which is bounded above by $\left(\frac{ek^2}{2(k+d)}\right)^{k+d}$. Then, for $k \leq d \leq \binom{k}{2} - k$, the above bound yields:

$$\begin{aligned}
C(k, d) &\leq k^{1/2} \left(\frac{e}{2}\right)^{k+d} d^{-d/2} k^{k+(3d-1)/2} \left(\frac{k}{k+d}\right)^{k+d/2} \left(\frac{d}{k+d}\right)^{d/2} \\
&< d^{-d/2} k^{k+(3d-1)/2} k^{1/2} \left(\frac{e}{2}\right)^{k+d} \exp\left(-(k+d/2)\frac{d}{k+d} - d/2\frac{k}{k+d}\right) \\
&< \sqrt{k} \left(\frac{\sqrt{e}}{2}\right)^{k+d} d^{-d/2} k^{k+(3d-1)/2}
\end{aligned} \tag{A.9}$$

, which is the same asymptotics as for $d \leq k$, as desired, as long as c for the previous theorem is at most 1, which Bollobás claims is true, but we won't need.

Appendix B

Other Deferred Proofs from the Erdős-Rényi Section

B.1 Reduction from $G(n, t)$ to $G(n, p)$.

Setting $p = 2t/n^2 = (1 + \epsilon)/n$, where $\epsilon = (2t/n) - 1 = 2s/n$, note that the probability of a particular k, d component existing in a $G(n, p)$ graph is bounded below by:

$$P_p = \sum_{t=0}^N \binom{N}{t} p^t q^{N-t} P_t \geq \binom{N}{M} p^M q^{N-M} P_M \geq P_M [e^{1/6M} (2\pi p q N)^{1/2}]^{-1} \quad (\text{B.1})$$

where P_t is the analogous probability of the event in $G(n, t)$, $M = \lfloor Nt \rfloor$, and the last argument following from a classic Stirling's approximation. By linearity of expectation, this implies $E_t(k) = \sum_d E_t[X(k, d)] \leq C \cdot [2\pi p q N]^{1/2} E_p(k)$. For our range of $p = 2t/n^2$, one can show that this implies $E_t(k) \leq 4n^{1/2} E_p(k)$: $((pqN)^{1/2} \approx t^{1/2} \approx n^{1/2})$

B.2 Proof of the Connectivity Result

Lemma 15. *The Erdős-Rényi graph process is almost surely connected by $p = \frac{4 \log n}{n}$.*

We shall defer the simple proof to the appendix.

Proof. While [35] gives a fascinating proof involving the spectral gap of the random graph Laplacian and the matrix Chernoff inequality, we shall stick to the more mundane proof.

The probability of the graph being disconnected is bounded above by the sum of the probabilities of there being a $1 \leq k \leq n/2$ isolated component, which is bounded above by:

$$\sum_{k=1}^{n/2} \binom{n}{k} (1-p)^{k(n-k)} \leq c \sum_{k=1}^{n/2} (en/k)^k e^{-pk(n-k)} \leq c \sum_{k=1}^{n/2} (en/k)^k e^{-pkn/2} \leq c \sum_{k=1}^{n/2} (e/n)^{-k} = o(1) \quad (\text{B.2})$$

□

Remark 4. *The result is far from optimal, with $p = \frac{\log n}{n}$ being the true threshold. This weak result is good enough for our purposes.*

This implies that for the Erdős Rényi graph process, the graph is almost surely connected by $t = n^2 p/2 = 2n \log n$.

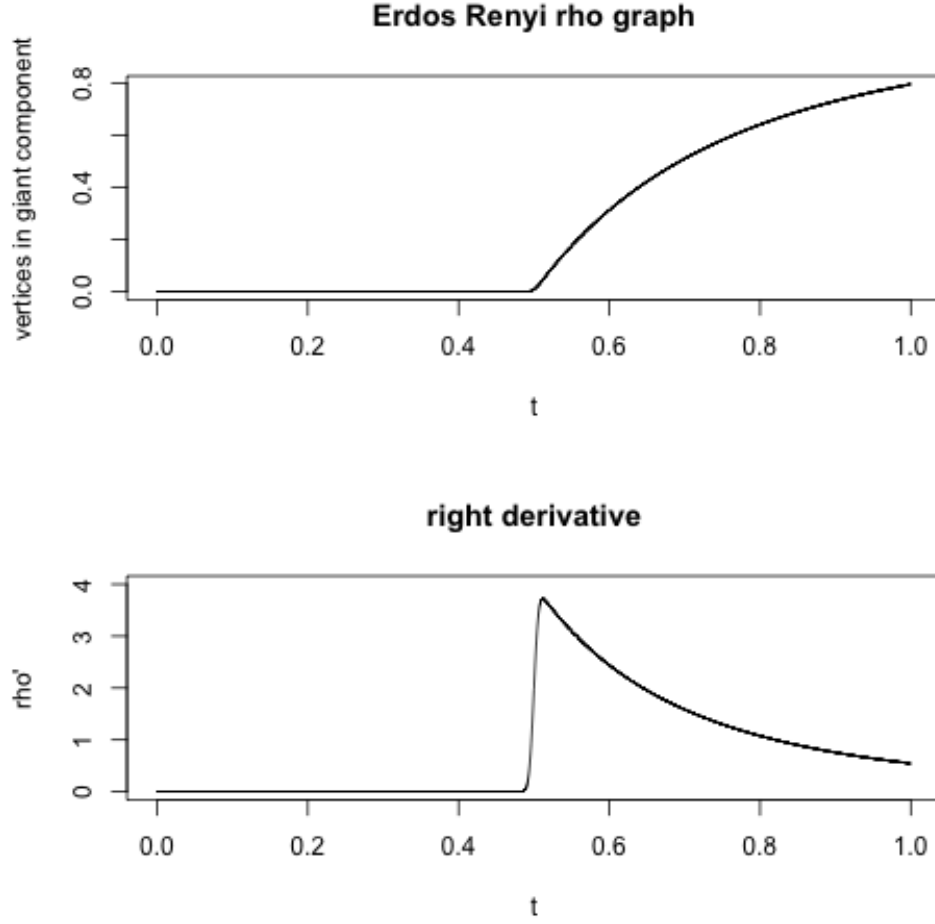


Figure B.1: Erdős Rényi Convergence Graph

B.3 Plot of ρ_{ER}

B.4 An alternative approach for the subcritical Erdős-Rényi graph: the random walk approach

[13] takes another approach to proving

$$P(L \leq \frac{1}{\alpha}(\log n - 5/2 \log \log n + \omega(n))) \mapsto 0 \quad (\text{B.3})$$

which involves no combinatorics, to obtain an almost optimal result. The key technique involves the random walk associated with the branching process approximation.

To give a rough idea of the proof, recall that the increments are $X_t = -1 + \text{Bin}(n, c/n)$. Intuitively, X_t represents picking one vertex, examining all of its neighbors, and then removing the original vertex from the vertices under consideration.

In fact, this argument can easily be made rigorous for the upper-bound. The idealized branching process stochastically dominates the actual exploration process, where the newly introduced neighbors are not fresh

draws from $\text{Bin}(n, c/n)$, but potentially smaller, as the neighbors can be from previously explored vertices.

Now, recall the martingale $S_t - (c-1)t$ and the stopping time of the process $T = \inf\{t : S_t = 0\}$. One can then compute the moment generating function $E[\exp(\theta X_i)]$, which is bounded above by $\exp(-\theta + c(e^\theta - 1)) = \psi(\theta)$, which is the moment generating function of the Poisson with mean c .

By optional stopping theorem for the supermartingale $M_t = \exp(\theta S_t)/\psi(\theta)^t$, one obtains $1/c = e^{\theta_1} \geq E[\psi(\theta_1)^{-T}] = E(e^{\alpha T})$. Then, Markov's inequality yields:

$$P(T \geq k) \leq e^{-k\alpha}/c \tag{B.4}$$

Hence, as $T \geq |C_x|$ in distribution, where C_x is the component size containing an arbitrary x , one can obtain the desired bound.

Appendix C

Sketch of the Proof for Sesqui-type Branching Process

We shall now sketch the results of [16], starting with an asymptotic formula for the probability $P(|\mathcal{S}| = N)$.

Theorem 40 (Asymptotic Point Probability of $\mathcal{S}$). *For $(Y^0, Z^0) \in \mathcal{K}^0$, $(Y, Z) \in \mathcal{K}$, and $|E[Y] - 1| \leq c_1$ with $c_1 > 0$ a constant, then for all $N \geq 1$, we have:*

$$P(|\mathcal{S}| = N) = N^{-3/2} e^{-N\xi} (\theta + O(N^{-1})) \quad (\text{C.1})$$

where $\xi = -\Psi(x^*) \geq 0$, $\theta = \sqrt{2\pi/|\Psi''(x^*)|} \Phi(x^*) = \Theta(1)$, and $\xi = \Theta(|EY - 1|^2)$.

The proof of this theorem is quite lengthy, laborious, and takes up almost 15 pages in the original paper. Hence, we shall not endeavour to give the whole proof, but rather give a sketch of the argument. Separating into particles of each type, $\mathcal{S}^L$ and $\mathcal{S}^S$, let $p_{n,m} = P(|\mathcal{S}^L| = n, |\mathcal{S}^S| = m)$.

The crucial step is to give an integral formula for $p_{n,m}$. As with the proof in the original branching process, we shall use the associated random walk of the branching process, where one selects one element from a list, examine its offsprings, and delete the element. Hence,

$$P(|\mathcal{S}^L| = n, |\mathcal{S}^S| = m | Y^0 = n_0, Z^0 = m_0) = P \left(n = \min_{n'} n_0 + \sum_{1 \leq j \leq n'} (Y_j - 1) = 0, m_0 + \sum_{1 \leq j \leq n} Z_j = m \right) \quad (\text{C.2})$$

To go from n being the first hitting time of 0 to a hitting time of 0, we need a result by Spitzer [30]:

Lemma 16. *Given a sequence $(y_1, \dots, y_n)$, $n_0 + \sum y_i = 0$, and $y_i \in \{-1, 0, 1, \dots\}$, there exists exactly n_0 cyclic permutations of $(y_1, \dots, y_n)$ such that n is the first hitting time of 0 and the cumulative sum staying positive before then.*

I shall omit the proof of this lemma, but intuitively consider the first hitting time of $n_0 - 1, n_0 - 2, \dots, 0$, which correspond to the n_0 cyclic shifts. One can argue that those are necessarily the only possible shifts. This technique has often been used to obtain results concerning the maxima of random walks, e.g. in [15].

Hence, using the exchangeability of (Y_i, Z_i) , we obtain:

$$P(|\mathcal{S}^L| = n, |\mathcal{S}^S| = m | Y^0 = n_0, Z^0 = m_0) = \frac{n_0}{n} P \left(n_0 + \sum_{1 \leq j \leq n'} Y_j = 0, m_0 + \sum_{1 \leq j \leq n} Z_j = m \right) \quad (\text{C.3})$$

Letting $g(Y, Z)$ be the PGF, the right-hand side probability is the $y^n z^m$ coefficient of $y^{n_0} z^{m_0} g(y, z)^n$. Denote $[y^n z^m]g$ to be the coefficient of $y^n z^m$ of g . Summing over n_0, m_0 , we obtain:

$$\begin{aligned} np_{n,m} &= \sum_{n_0, m_0} P(Y^0 = n_0, Z^0 = m_0) n_0 [y^n z^m] (y^{n_0} z^{m_0} g(y, z)^n) \\ &= [y^n z^m] \sum_{n_0, m_0} P(Y^0 = n_0, Z^0 = m_0) n_0 y^{n_0} z^{m_0} g(y, z)^n \\ &= [y^n z^m] \left(y \frac{\partial}{\partial y} g_{Y^0, Z^0}(y, z) g(y, z)^n \right) \end{aligned} \quad (\text{C.4})$$

Denote $y \frac{\partial}{\partial y} g_{Y^0, Z^0}$ as $\tilde{g}_0$. Also, denote $f(y, z) = E[e^{yY+zZ}] = g(e^y, e^z)$ to be the moment generating function (MGF) and $\tilde{f} = \frac{\partial}{\partial y} f(y, z) = E[Y e^{yY+zZ}]$, and $f_0, \tilde{f}_0$ analogously.

Quite beautifully, as $\tilde{g}_0, g$ are analytic, one can use Cauchy's integral formula to obtain the $y^n z^m$ coefficient:

$$np_{n,m} = \frac{1}{(2\pi i)^2} \oint \oint y^{-n} z^{-m} \tilde{g}_0(y, z) g(y, z)^n \frac{dy}{y} \frac{dz}{z} \quad (\text{C.5})$$

where the integral is defined over circles of center 0 and radii such that $\tilde{g}$ and g are defined. If we assume $(Y, Z), (Y^0, Z^0) \in \mathcal{K}^0$, then for $\alpha, \beta < \log R$, one can integrate over $|y| = e^\alpha, |z| = e^\beta$ to obtain:

$$np_{n,m} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{-n(\alpha+iu)-m(\beta+iv)} \tilde{f}_0(\alpha+iu, \beta+iv) f(\alpha+iu, \beta+iv)^n du dv \quad (\text{C.6})$$

The rest of the proof of Theorem 40 involves estimating this integral asymptotically, exploiting the assumption that $(Y, Z) \in \mathcal{K}^0$. The core of the argument involves estimating the exponential integral using a two-dimensional generalization of Laplace's method (also called the saddle-point method), which replaces the original integral with a Gaussian integral. Under an opportune choice of α, β , one can thus obtain:

$$p_{n,m} = n^{-2} e^{n\psi(\alpha, \beta)} \left((2\pi)^{-1} \tilde{f}_0(\alpha, \beta) \text{Det}(D^2\phi(\alpha, \beta))^{-1/2} + O(n^{-1}) \right) \quad (\text{C.7})$$

, where $\phi(y, z) = \log f(y, z)$ is an analytic extension of the cumulant function, and ψ the residual term after linearizing ϕ at (α, β) , i.e. $\psi(\alpha, \beta) = \phi(\alpha, \beta) - \alpha D_1\phi(\alpha, \beta) - \beta D_2\phi(\alpha, \beta)$.

Finally, to compute $P(|\mathcal{S}| = N) = \sum_{n=0}^N p_{n, N-n}$, as an individual of type L has on average $E[Y]$ children of type L and $E[Z]$ children of type S . Hence, for $E[Y] \approx 1$, the overall fraction of type L individuals in $\mathcal{S}$ should be close to $x_0 = 1/(1 + EZ)$. Thus, one can show that $p_{n, N-n}$, with $n/N \approx x_0$ dominate the sum. Adding the terms near $n/N \approx 1/(1 + EZ)$ and bounding away the rest, one is able to obtain Theorem 40.

Appendix D

Proof of Relevant Inequalities: Hoeffding-Azuma, McDiarmid, and Chernoff

Theorem 41 (Hoeffding-Azuma's Inequality). *Suppose X_n is a martingale with respect to a filtration F_n satisfying the bounded differences condition with $P(|X_n - X_{n-1}| \leq c_n) = 1$. Then, for any $n > 0$, we have:*

$$P(|X_n - X_0| > t) \leq 2 \exp\left(-\frac{t^2}{2 \sum_{i=1}^n c_i^2}\right) \quad (\text{D.1})$$

Proof. By Jensen, for $|x/c| < 1$, we can express $\exp(\lambda x)$ as follows:

$$\begin{aligned} \exp(\lambda x) &= \exp\left(\lambda \left(\frac{1}{2}\left(1 + \frac{x}{c}\right)c + \frac{1}{2}\left(1 - \frac{x}{c}\right)(-c)\right)\right) \leq \frac{1}{2}\left(1 + \frac{x}{c}\right) \exp(\lambda c) + \frac{1}{2}\left(1 - \frac{x}{c}\right) \exp(-\lambda c) \\ &= \cosh(\lambda c) + \frac{x}{c} \sinh \lambda c \end{aligned} \quad (\text{D.2})$$

From the Taylor series, one can easily bound $\cosh(x) \leq \exp(\frac{x^2}{2})$, and hence from the above, we obtain for $|x| < c$:

$$\exp(\lambda x) \leq \exp\left(\frac{\lambda^2 c^2}{2}\right) + \frac{x}{c} \sinh \lambda c \quad (\text{D.3})$$

Turning to the original inequality, we have:

$$\begin{aligned} P(X_n - X_0 > t) &= P(\exp(X_n - X_0) > e^t) \leq \exp(-\lambda t) E\left[\exp\left(\sum_{i=1}^n \lambda(X_i - X_{i-1})\right)\right] \\ &= \exp(-\lambda t) E\left[\exp\left(\sum_{i=1}^{n-1} \lambda(X_i - X_{i-1})\right) E[\exp(\lambda(X_n - X_{n-1})) | F_{n-1}]\right] \end{aligned} \quad (\text{D.4})$$

where the last line follows by iterated expectations. Applying the inequality above, one can replace the exponent term by $\exp(\frac{\lambda^2 c_n^2}{2}) + C(X_n - X_{n-1})$, and as X is a martingale, the linear term vanishes under the expectation. Hence, by repeated iteration, one obtains:

$$P(X_n - X_0 > t) \leq \exp(-\lambda t) \exp\left(\frac{\sum \lambda^2 c_i^2}{2}\right) \quad (\text{D.5})$$

Setting $\lambda = \frac{t}{\sum_{i=1}^n c_i^2}$ gives us $P(X_n - X_0 > t) \leq \exp(-\frac{t^2}{2 \sum_{i=1}^n c_i^2})$. An identical argument will bound $P(X_n - X_0 < -t)$, giving us the inequality. $\square$

Corollary 14. *The exact argument can be made for X_i being supermartingales. For X_i a submartingale, as $-X_i$ is a supermartingale, we obtain upper bounds on the lower tails of submartingales as well.*

McDiarmid's Inequality follows largely as a corollary.

Corollary 15 (McDiarmid's Inequality). *Suppose $f : \mathbb{R}^n \mapsto \mathbb{R}$ is Lipschitz with respect to the Hamming metric, i.e. if x and x' only differ in the k -th coordinate, there exists $\sigma_k > 0$ such that $|f(x) - f(x')| \leq \sigma_k$. Then, given $Y = f(X_1, \dots, X_n)$ for X_i independent and any $\alpha > 0$:*

$$P(|Y - EY| \geq \alpha) \leq 2 \exp\left(-\frac{2\alpha^2}{\sum \sigma_i^2}\right) \quad (\text{D.6})$$

Proof. Simply let $Y_k = E[Y|X_1, X_2, \dots, X_k]$, which is a martingale, and by our definition of f , satisfies $|Y_k - Y_{k-1}| = |E[Y - Y'|X_1, \dots, X_k]|$, where $Y' = f(X'_1, \dots, X'_n)$, where X' is the same as X save for the k -th coordinate, which is drawn independently. Thus, by our restriction on f , $|Y_k - Y_{k-1}| \leq \sigma_k$, so we obtain the conditions of Azuma-Hoeffding. $\square$

Our final helpful inequality is the frequently used Chernoff lemma.

Lemma 17 (Chernoff-Bounds). *Let X_i , $1 \leq i \leq M$, be a sequence of Bernoulli random variables. Let $P(X_j|X_{1:j-1}) > q$. Then, there is an absolute constant $\eta > 0$ such that:*

$$P\left(\sum_{i=1}^M X_i < qM/2\right) < \exp(-\eta qM) \quad (\text{D.7})$$

Proof.

$$P\left(\sum_{i=1}^M X_i < qM/2\right) = P(\exp((-t) \sum X_i) > \exp(-tqM/2)) < E[\exp((-t) \sum X_i)] \exp(tqM/2) \quad (\text{D.8})$$

Next, as $qe^{-t} + (1-q) < \exp((e^{-t} - 1)q)$ for all $q \in (0, 1)$, by the law of iterated expectation, one obtains:

$$P\left(\sum_{i=1}^M X_i < qM/2\right) < \exp\left(qM(e^{-t} - 1 + \frac{1}{2}t)\right) \quad (\text{D.9})$$

As $e^{-t} \approx 1 - t$ as t gets small, for a sufficiently small t , there is an absolute constant $\eta > 0$ such that:

$$P\left(\sum_{i=1}^M X_i < qM/2\right) < \exp(-\eta qM) \quad (\text{D.10})$$

$\square$

Appendix E

Proof that $\mathcal{S}_t$ is a t_c -critical branching process family

In this section, we shall give a proof that the branching process family $\mathcal{S}_t$ is indeed a t_c -critical branching process family. Recall the definition:

Definition 12. *The branching process family $(\mathcal{S}_{Y_t, Z_t, Y_t^0, Z_t^0})$ is t_c -critical if the following hold:*

1. *There exist $\delta > 0$, $R > 1$ with $(t_c - \delta, t_c + \delta) \subset (t_0, t_1)$ such that the generating functions:*

$$g(t, \alpha, \beta) = E[\alpha^{Y_t} \beta^{Z_t}] \text{ and } g^0(t, \alpha, \beta) = E[\alpha^{Y_t^0} \beta^{Z_t^0}] \quad (\text{E.1})$$

are defined for all real t with $|t - t_c| < \delta$ and all complex α, β , $|\alpha|, |\beta| < R$. Furthermore, these functions have analytic extensions to t now complex and $|t - t_c| < \delta$.

- 2.

$$EY_{t_c} = 1, EY_{t_c}^0 > 0, \text{ and } \frac{d}{dt} EY_t|_{t=t_c} > 0 \quad (\text{E.2})$$

3. *There exist some $k_0 \in \mathbb{N}$ such that*

$$\min \{P(Y_{t_c} = k_0, Z_{t_c} = k_0), P(Y_{t_c} = k_0 + 1, Z_{t_c} = k_0), P(Y_{t_c} = k_0, Z_{t_c} = k_0 + 1)\} > 0 \quad (\text{E.3})$$

Hence, we shall verify each of the three prerequisites, starting with the analyticity condition.

Theorem 42. *There exist $\delta > 0$ and $R > 1$ such that $g(t, \alpha, \beta) = E[\alpha^{Y_t} \beta^{Z_t}]$ and $g^0(t, \alpha, \beta) = E[\alpha^{Y_t^0} \beta^{Z_t^0}]$ are defined for all real $|t - t_c| < \delta$ and complex α, β with $|\alpha|, |\beta| < R$. Furthermore, each of these functions has an analytic extension to $\{(t, \alpha, \beta) : |t - t_c| < \delta, |\alpha|, |\beta| < R\}$.*

Proof. Here, we shall cite the following critical lemma, which borrows results from PDE theory:

Lemma 18. *Define the generating from for $q_{k,r}(t)$:*

$$P(t, x, y) = \sum_{k,r \geq 0} x^k y^r q_{k,r}(t) \quad (\text{E.4})$$

Then, P is real analytic on $(t_0, t_1) \times (-C, C) \times (-C, C) \in \mathbb{R}^3$ for a constant C . Furthermore, for each $t \in (t_0, t_1)$, there exists a neighborhood δ such that there exists an analytic extension to $|t' - t| < \delta$, $|x|, |y| < C$.

Proof. We shall just give a sketch of the proof. Recall that we have obtained the subexponential bound: $\sup_t q_{k,r}(t) \leq B e^{-b(k+r)}$. Hence, the above series converges absolutely.

Next, one can use the differential equations obtained for q' in equation 6.28, to obtain a partial differential equation (q' indicates a first order partial with respect to time t) satisfied by $P(t, x, y)$.

Then, the Cauchy-Kovalevskaya theorem gives us a local analytic solution $\tilde{P}$. To show that $\tilde{P}$ and P coincide (if both are analytic, they trivially do, but in general they may differ by a non-analytic solution), one can work backwards to obtain $\tilde{q}$, i.e. the coefficients for $\tilde{P}$, and show that they satisfy the same differential equations as q , and must coincide. $\square$

Now, given an analytic extension $P(t, x, y) = \sum_{k,r \geq 0} x^k y^r q_{k,r}(t)$, the result follows easily. First, let us analyze the generating function for N_k :

$$\Phi(\alpha) = \frac{1}{\rho_\omega(t_0)} \sum_{k \geq 0} \alpha^k \rho_k(t_0) \quad (\text{E.5})$$

Recall the exponential tail bound $|\rho_k(t_0)| < A e^{-ak}$, which implies that $\Phi(\alpha)$ is analytic for all $|\alpha| < e^a$. Furthermore, as $\Phi(1) = 1$ and all the terms in the power series is positive, one can pick $\alpha_0 \in (1, e^a)$ such that $|\Phi(\alpha)| < e^{b/3} = \beta_0$ for $|\alpha| < \alpha_0$.

The generating function for (Y_t, Z_t) is given by:

$$g(t, \alpha, \beta) = \prod_{k \geq 0, r \geq 1} E \left(E(\alpha^N)^{r-1} \beta^k \right)^{H_{k,r,t}} \quad (\text{E.6})$$

As $H_{k,r,t} = Po(rq_{k,r}(t)/\rho_\omega(t_0))$, by using the Poisson generating function, we obtain:

$$\begin{aligned} g(t, \alpha, \beta) &= \exp \left(\sum_{k \geq 0, r \geq 1} (rq_{k,r}(t) ((\Phi(\alpha))^{r-1} \beta^k - 1) / \rho_\omega(t_0)) \right) \\ &= \exp ((P_y(t, \beta, \Phi(\alpha)) - P_y(t, 1, 1)) / \rho_\omega(t_0)) \end{aligned} \quad (\text{E.7})$$

Now, the result follows from the analytic extension of P_y , as desired.

The result for g^0 on (Y^0, Z^0) follows analogously. $\square$

Corollary 16. *As $g_\alpha(t, 1, 1) = EY_t$, $g_{\alpha\alpha}(t, 1, 1) = EY_t(Y_t - 1)$, we obtain that EY_t , EY_t^2 , $VarY_t$ are all analytic for $t \in (t_c - \epsilon, t_c + \epsilon)$.*

Thus, suffices to show the remaining two conditions to conclude that $\mathcal{S}_t$ is t_c -critical. First, by our parity assumption, the third condition, there exists a k_0 such that:

$$\min \{P(Y_{t_c} = k_0, Z_{t_c} = k_0), P(Y_{t_c} = k_0 + 1, Z_{t_c} = k_0), P(Y_{t_c} = k_0, Z_{t_c} = k_0 + 1)\} > 0 \quad (\text{E.8})$$

holds trivially for $k_0 = 1$. Hence, suffices to demonstrate the second prerequisite, concerning EY_t .

Theorem 43. *$EY_{t_c} = 1$, and for all $t \in (t_0, t_1)$, we have:*

$$EY_t^0 > 0 \text{ and } \frac{d}{dt} EY_t > 0 \quad (\text{E.9})$$

Proof.

$$EY_t = \sum_{k \geq 0, r \geq 2} EH_{k,r,t} \cdot (r-1) \cdot EN = EN / \rho_\omega(t_0) \cdot \left(\sum_{k,r \geq 0} r(r-1) q_{k,r}(t) \right) \quad (\text{E.10})$$

Note that $q_{0,2}$ is increasing: it is an idealized process for the number of $V_L - V_L$ edges added up to time t , and hence increases. Similarly, to show the rest of the terms to be increasing in t , suffices to show that the random quantity $W = \sum_{k \geq 1, r \geq 2} r(r-1) Q_{k,r}(i)$ does not decrease in i . This can be easily verified by checking the $V_S - V_L$ or $V_S - \bar{V}_S$ edges.

As for $EY^0 > 0$, we simply have:

$$EY_t^0 \geq EY_{0,t} = EN \cdot \rho_\omega(t_0) > 0 \quad (\text{E.11})$$

Finally, to show $EY_{t_c} = 1$, note that $\rho(t) = P(|\mathcal{S}_t| = \infty)$, and hence $P(|\mathcal{S}_t|) = 0$ for $t \in [t_0, t_c]$, and $P(|\mathcal{S}_t|) > 0$ for $t \in (t_c, t_1]$. Note that the branching process $\mathcal{S}$ survives when $EY_t > 1$, goes extinct when $EY_t < 1$, and is indeterminate at $EY_t = 1$, so the above result shows $EY_t < 1$ for $t < t_c$, $EY_t > 1$ for $t \geq t_c$, and as EY_t is analytic in t , $EY_{t_c} = 1$, as desired. $\square$

Hence, we have shown the branching process $\mathcal{S}_t$ is indeed a t_c -critical branching process family. Hence, we have obtained:

Theorem 44. $\mathcal{S}_t$ the branching process coupled with the random graph exploration process, satisfies the following:

1. there exist constants $\epsilon_0, c > 0$ and analytic functions θ and ψ on $[t_c - \epsilon_0, t_c + \epsilon_0]$ such that:

$$P(|\mathcal{S}| = k) = (1 + O(1/k))k^{-3/2}\theta(t)e^{-k\psi(t)} \quad (\text{E.12})$$

with $\theta > 0$, $\psi \geq 0$, with $\psi(t_c) = \psi'(t_c) = 0$ and $\psi''(t_c) > 0$.

2. The survival probability $\rho(t) = P(|\mathcal{S}_t| = \infty) = 0$ for $t_c - \epsilon \leq t \leq t_c$, and is positive for $t_c < t \leq t_c + \epsilon_0$.
3. $\rho(t)$ is analytic on $[t_c, t_c + \epsilon_0]$.