The phase transition in the Erdös-Rényi random graph model

Tom van den Bosch

July 17, 2014

Bachelor thesis
Supervisor: Guus Regts

Korteweg-de Vries Instituut voor Wiskunde
Faculteit der Natuurwetenschappen, Wiskunde en Informatica
Universiteit van Amsterdam
Abstract

In this thesis we intensively study the phase transition in the Erdös-Rényi random graph model. First, the definition of random graphs is given. After that we show a proof of a classical theory by Erdös by using the probabilistic method. With the probabilistic method we will study the Erdös-Rényi random graph model around $p = \frac{1}{n}$, where a giant component emerges. An elaborate study of this phase transition is given. At last we will give examples of some other random graph models and prove that the phase transition in the random graph models is continuous.

Title: The phase transition in the Erdös-Rényi random graph model
Authors: Tom van den Bosch, tom.vd.bosch@gmail.com, 10196579
Supervisor: Guus Regts
Second grader: Prof. Bernard Nienhuis
Date: July 17, 2014

Korteweg-de Vries Instituut voor Wiskunde
Universiteit van Amsterdam
Science Park 904, 1098 XH Amsterdam
http://www.science.uva.nl/math
Contents

1. Introduction 4
   1.1. Introduction to random graphs 4
   1.2. Definitions 5
   1.3. Acknowledgments 6

2. Erdös’ Theorem 7

3. The phase transition 10
   3.1. Introduction to the phase transition 10
   3.2. Breadth First Search algorithm and the Galton-Watson branching process 12
      3.2.1. Breadth First Search 12
      3.2.2. The Galton-Watson branching process 13
   3.3. Three branching models 14
      3.3.1. The graph branching process 14
      3.3.2. The binomial branching model 14
      3.3.3. The Poisson branching model 14
   3.4. analyzing the branching models 15
      3.4.1. The graph branching model 15
      3.4.2. Linking the models 16
   3.5. The subcritical regimes 20
      3.5.1. The very subcritical phase 20
      3.5.2. The barely subcritical phase 21
   3.6. The supercritical regimes 22
      3.6.1. The very supercritical phase 22
      3.6.2. The barely supercritical phase 23
   3.7. The critical window 24

4. Other random graph models and continuity 27
   4.1. Some other random graph models 27
   4.2. Continuity of the phase transition 28

5. Popular summary 31

Bibliography 33

A. Probability distributions and Chernoff bounds 35
1. Introduction

1.1. Introduction to random graphs

Random graphs were first introduced by Paul Erdős and Alfréd Rényi in 1959 and independently in that same year by Edgar Gilbert. Both introduced models similar to one another, and we now know both of them as the Erdős-Rényi random graph model. These models can be used to prove that properties hold for almost all graphs (meaning there is a countable number of graphs not having that property). They can also be used to prove the existence of graphs with certain properties with a method called the probabilistic method, which we will show a classical example of in chapter 2.

Applications of random graphs can be found in network theory (see Newman 2002 [1]) as any network can be represented by a graph. Another field in which random graphs are useful is percolation theory. Alon and Spencer [5] in The probabilistic method have devoted a section to compare random graphs to bond percolation, which has been studied intensively by mathematicians and physicists since it was introduced by Broadbent and Hammersley in 1957 [2].

Before defining random graphs, it is important to know that a random graph is not a single graph, but rather a probability space on all possible edge combinations of a vertex set $V$ of $n$ vertices. In random graphs the number of vertices $n$ usually tends to infinity.

There are two models we use to look at random graphs, both very closely related. These two models are called the Erdős-Rényi random graph models. The first model was introduced by Gilbert and is called $G(n, p)$, which includes any edge $e$ in the edge set $V^2$ with probability $p$ in the random graph. The second model is the $G(n, m)$ model, which uniformly takes a graph from all graphs with $n$ vertices and $m$ edges. Another way to see the $G(n, m)$ model is to start with an empty graph on $n$ vertices and add $m$ edges from $V^2$ at random to the graph. We will mostly use the $G(n, p)$ model. The models $G(n, p)$ and $G(n, m)$ behave similarly for $m = \binom{n}{2} p$, given that $n$ is very large. Even though the models behave similarly, they are not exactly the same. For example, the probability that a graph $G \in G(n, p)$ has an even number of edges is some number between 0 and 1, while that same probability for $G(n, m)$ is always exactly 0 or 1.

One of the most interesting things of random graphs is the Erdős-Rényi Phase transition. This is the interval in the random graph model where the size of the largest component shows very rapid growth. Think of it like the empire of Genghis Khan. At first, there were many small countries. When one of them started conquering the other countries, it grew bigger and kept growing bigger, while all other countries that were not conquered stayed small. Of course, the Mongolian empire had to fall one day and random graphs just keep on growing until we have a complete graph.
In the $\mathcal{G}(n, p)$ model this phase transition is around $p = \frac{1}{n}$. If $p$ is smaller than that, all components of the random graph will have size of order at most $\ln n$. If $p > \frac{1}{n}$, a giant component of size of order $n$ emerges, while all smaller components are of order $\ln n$. An elaborate study of this phase transition will be done in chapter 3.

More recently mathematicians have studied the continuity of the phase transition. The continuity of the phase transition means that the size of the largest component of the random graph is continuously dependent on the number of edges. Aside from the continuity of the phase transition in the Erdős-Rényi random graph model, there are also other random graph models which have such a phase transition that is continuous. We will be showing these models and the continuity of the phase transition in chapter 4.

1.2. Definitions

Some of the definitions used in chapters 2 and 3 might not be native to graph theory, so they will be given here.

First there are two similar terms we used to describe an event that has probability (tending to) one:

- Whp: With high probability. An event (dependent on $n$) occurs with high probability if for any $\alpha \geq 1$ the event occurs with probability at least $1 - \frac{c_\alpha}{n^\alpha}$, where $c_\alpha$ depends only on $\alpha$. This means that the probability with $n \to \infty$ goes to one.

- A.s.: Almost surely. An event happens almost surely if it happens with probability one, usually meaning there is an uncountable infinite set of possibilities and the set of possibilities in which the event does not occur is countable.

Then there are the terms we used to describe that something is relatively small compared to other terms:

- $O(n)$: Let $f$ and $g$ be two functions defined on some subset of the real numbers. Then $f(x) = O(g(x))$ if and only if there is a positive constant $M$ and a positive real $x_0$ such that for all sufficiently large $x \geq x_0$, $|f(x)| \leq M|g(x)|$. So if $C(v) = O(\ln n)$ this means that for a constant $c$, $C(v) \leq c \ln n$ for all $n$ sufficiently large. The case where something is $O(1)$ means it is a constant.

- $o(n)$: Let $f$ and $g$ be two functions defined on some subset of the real numbers. Then $f(x) = o(g(x))$ as $n \to \infty$ if and only if for every real $\epsilon > 0$ there exists a constant $x_0$ such that $|f(x)| \leq \epsilon |g(x)|$ for all $x \geq x_0$. Note that this is a stronger statement than $O(n)$ as for $O(n)$ the statement has to be true for at least one constant $M$, but for $o(n)$ it has to be true for all $\epsilon > 0$. Every function $f$ that is $o(g)$ is also $O(g)$, but the inverse is not true. For example, $g \neq o(g)$.

- $\Omega(n)$: Let $f$ and $g$ be two functions defined on some subset of the real numbers. Then $f(x) = \Omega(g(x))$ for $x \to \infty$ if and only if $\limsup_{x \to \infty} \frac{|f(x)|}{|g(x)|} > 0$. The only case
we used this is when saying that there are $\Omega(n^2)$ pairs of vertices, meaning the amount of pairs of vertices is at least a constant times $n^2$.

1.3. Acknowledgments

I would like to thanks Guus Regts for being the supervisor for my bachelor thesis. His understanding of this subject helped me tremendously in understanding random graphs, and he helped me in writing mathematics in general. He also provided sources for me to study, and insights on subjects that those sources where not clear on. Second, I would like to thanks Bernard Nienhuis for being my second grader.
2. Erdős’ Theorem

Now that we know the notion of a random graph, we will be giving an interesting proof involving random graphs. The girth of a graph is length of the smallest cycle in the graph. If there is no cycle in a graph its girth is infinity. A theorem by Erdős says that for any number \( k \) there exists a graph with girth at least \( k \) and chromatic number at least \( k \). This theorem might seem counterintuitive, as large girth means that locally a graph looks like a tree and thus is two-colorable. In this chapter we will prove this theorem with the probabilistic method. This is a nonconstructive method used to prove the existence of a mathematical object with certain properties, pioneered by Erdős. All proofs and results in this chapter are from Diestel (2010) [3].

Before we can prove Erdős’ theorem, we need some lemmas. A set of \( k \) vertices is called independent if there are no edges between any of the vertices in the set. The clique number \( \omega(G) \) is the largest integer \( k \) for which the graph \( G \) has a subgraph \( K^k \). The independence number \( \alpha(G) \) is the largest integer \( k \) for which \( G \) has a set of \( k \) independent vertices.

**Lemma 2.1.** For all integers \( n,k \) with \( n \geq k \geq 2 \), the probability that \( G \in \mathcal{G}(n,p) \) has a set of \( k \) independent vertices is at most

\[
\Pr[\alpha(G) \geq k] \leq \binom{n}{k} (1 - p)^{\binom{k}{2}}.
\]

And also the probability that \( G \in \mathcal{G}(n,p) \) contains a \( k \)-set that is a complete subgraph \( K^k \) has the bound

\[
\Pr[\omega(G) \geq k] \leq \binom{n}{k} (p)^{\binom{k}{2}}.
\]

**Proof.** The probability for a fixed \( k \)-set of to be independent in \( G \) is \((1 - p)^{\binom{k}{2}}\). As there are \( \binom{n}{k} \) \( k \)-sets in \( V \), this gives us the desired inequality. For the second claim the proof is exactly the same if we replace \( 1 - p \) by \( p \). \( \square \)

Even though we technically have our proof of the probability not being greater than \( \binom{n}{k} (1 - p)^{\binom{k}{2}} \), we want to know that its not always an equality. Let’s take a look at the case with \( n = 3 \) and \( k = 2 \). Our bound gives \( \binom{3}{2} (1 - p)^{\binom{2}{2}} = 3 (1 - p) \). However, the only graph on 3 vertices without 2 independent vertices is the complete graph \( K^3 \). The probability of \( G \in \mathcal{G}(3,p) \) being \( K^3 \) is only \( p^3 \), so the probability of there being 2 independent vertices is \( 1 - p^3 \). For \( p \in [0, 1] \) we know \( 1 - p^3 \leq 3(1 - p) \) (we have equality only at \( p = 1 \)), so the bound is not exact. Again replacing \( 1 - p \) by \( p \) gives us the same for the second claim.
The next lemma tells us about the number of \( k \)-cycles in a random graph. A \( k \)-cycle is, as expected, a cycle with length exactly \( k \).

**Lemma 2.2.** The expected number of \( k \)-cycles in \( G(n, p) \) is \( \frac{n!}{(n-k)!2^k}p^k \).

*Proof.* There are \( n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!} \) ways to choose \( k \) distinct vertices in \( V \). However, there are \( 2k \) ways of picking the same cycle, as we can start picking vertices at any of the \( k \) vertices in the cycle, and we can pick the vertices going either clockwise or counterclockwise in the cycle. Thus there are \( \frac{n!}{(n-k)!2^k} \) possible cycles in \( G \). Since there are \( p^k \) edges in a cycle, the probability for a fixed cycle to be in \( G \) is \( p^k \). If we sum this over all possible cycles we get the desired \( \frac{n!}{(n-k)!2^k}p^k \). \( \square \)

The next lemma is a classical result from probability theory. It is called Markov’s inequality, though it is sometimes also called Chebyshev’s first inequality, as it first appeared in the works of Markov’s teacher, Chebyshev.

**Lemma 2.3** (Markov’s inequality). Let \( X \geq 0 \) be a random variable on \( G(n, p) \) and \( a > 0 \), then

\[
\Pr[X \geq a] \leq \frac{1}{a}E(X).
\]

*Proof.* Simply using the definition of the expectation we get

\[
E(X) = \sum_{G \in \mathcal{G}(n, p)} \Pr\{G\} \cdot X(G) \geq \sum_{G \in \mathcal{G}(n, p), \ X(G) \geq a} \Pr\{G\} \cdot a = \Pr[X \geq a] \cdot a.
\]

Dividing by \( a \) on both sides yields the desired result. \( \square \)

As Markov’s inequality is also called Chebyshev’s first inequality, there must be Chebyshev’s second inequality, which we will show below. First though, we need the definition of the *standard deviation* used in the lemma: \( \sigma^2 = E((X - \mu)^2) \). Now we can proof the following theorem.

**Theorem 2.4** (Chebyshev’s inequality). For all real \( \lambda > 0 \)

\[
\Pr[|X - \mu| \geq \lambda] \leq \frac{\sigma^2}{\lambda^2}.
\]

*Proof.* By Markov’s inequality and the definition of \( \sigma^2 \)

\[
\Pr[|X - \mu| \geq \lambda] = \Pr[(X - \mu)^2 \geq \lambda^2] \leq \frac{1}{\lambda^2}E(X - \mu)^2 = \frac{\sigma^2}{\lambda^2}.
\]

\( \square \)

Chebyshev’s inequality will not be used for proving Erdös’ theorem, but it is an important theorem in probability theory.

We now need just one last lemma to be able to prove Erdös’ theorem.
Lemma 2.5. Let $k \in \mathbb{N}$ and let $p = p(n)$ be a function of $n$ such that $p \geq (6k \ln n)n^{-1}$ for large $n$. Then $\lim_{n \to \infty} \Pr[\alpha \geq \frac{n}{2k}] \to 0$ in $G(n, p)$.

Proof. For all integers $n, r$ with $n \geq r \geq 2$ and all $G \in G(n, p)$ Lemma 2.1 implies

$$\Pr[\alpha \geq r] \leq \binom{n}{r}(1-p)^{(\frac{r}{2})} \leq n^r(1-p)^{(\frac{r}{2})} = (n(1-p)^{\frac{r}{2}})^r \leq (ne^{-\frac{1}{2}p(r-1)})^r.$$

Where the last inequality follows from the fact that $1-p \leq e^{-p}$ for all $p$. Now if $p \geq (6k \ln n)n^{-1}$ and $r \geq \frac{n}{2k}$ the term under the exponent satisfies

$$ne^{-\frac{1}{2}p(r-1)} = ne^{-\frac{1}{2}p(\frac{r}{2})} \leq ne^{-\frac{3}{4} \ln n + \frac{1}{2}p} \leq nn^{-\frac{3}{2}e^\frac{1}{2}} = \left(\frac{e}{n}\right)^{\frac{1}{2}}.$$

This last expression goes to 0 when $n \to \infty$. Thus for $r := \left\lceil \frac{n}{2k} \right\rceil$ we obtain

$$\lim_{n \to \infty} \Pr[\alpha \geq \frac{n}{2k}] = \lim_{n \to \infty} \Pr[\alpha \geq r] = 0$$

as claimed. \qed

Now we can prove the theorem that we have been hyping up for this whole chapter.

Theorem 2.6 (Erdős 1959). For every integer $k$ there exists a graph $H$ with girth $g(H) > k$ and chromatic number $\chi(H) > k$.

Proof. We assume $k \geq 3$ and fix $\epsilon$ with $0 < \epsilon < \frac{1}{k}$. Let $p := n^{-\epsilon}$. Let $X$ denote the number of cycles in a random graph with length at most $k$, called short cycles. By Lemma 2.2 we have

$$\mathbb{E}(X) = \sum_{i=3}^{k} \frac{n!}{(n-i)!2i!}p^i \leq \frac{1}{2} \sum_{i=3}^{k} n^ip^i \leq \frac{1}{2}(k-2)n^kp^k.$$

Where $n^ip^i \leq n^kp^k$ because $np = n^\epsilon \geq 1$. Now by Markov’s inequality

$$\Pr[X \geq \frac{n}{2}] \leq \frac{2}{n} \mathbb{E}(X) \leq (k-2)n^{k-1}p^k \leq (k-2)n^{k-1}n^{(\epsilon-1)k} = (k-2)n^{k\epsilon-1}.$$

Because $k\epsilon - 1 < 0$ this implies $\lim_{n \to \infty} \Pr[X \geq \frac{n}{2}] = 0$. Now let $n$ be large enough such that $\Pr[X \geq \frac{n}{2}] < \frac{1}{2}$ and $\Pr[\alpha \geq \frac{n}{2k}] < \frac{1}{2}$. This is possible because of Lemma 2.5 and our choice of $p$. Then there is a graph $G \in G(n, p)$ with fewer than $\frac{n}{k}$ short cycles and $\alpha(G) < \frac{n}{2k}$. Let $H$ be the graph that is obtained from $G$ by deleting a vertex from every short cycle. Then $|H| \geq \frac{n}{2}$ and there are no short cycles in $H$, so $g(H) > k$. By definition of $G$

$$\chi(H) \geq \frac{|H|}{\alpha(H)} \geq \frac{n}{\alpha(G)} > k.$$

\qed

Erdős theorem is one of the earliest results of the probabilistic method. We will be using the probabilistic method more in chapter 3.
3. The phase transition

3.1. Introduction to the phase transition

In their work on the evolution of random graphs (1960) [4], Erdős and Rényi expressed interest in the random graph $G(n, m)$ where $m$ was near $\frac{1}{2}n$. They made the interesting discovery that around $m = \frac{1}{2}n$ the size of the largest component grows very rapidly. More specifically, for $m < \frac{1}{2}n$ the largest component of $G(n, m)$ is of the order $\ln n$, and for $m > \frac{1}{2}m$ it grows to order $n$. This rapid growth is called the Erdős-Rényi phase transition. Since that discovery, many studies have been done on this subject. In this chapter we will be looking at the phase transition more closely, as well as showing that the rapid growth of the largest component is continuous, even though some apparently thought it was not. Lastly, we will be showing some other random graph models that delay the phase transition. All proofs and results from this chapter, except where noted, are from Alon & Spencer (2008) [5]

Although Erdős and Rényi used the $G(n, m)$ model, we will be studying the $G(n, p)$ model, as is more common in modern studies. The phase transition then occurs at $p = \frac{1}{n}$. We will be looking at five different intervals of the phase transition, called the very subcritical, barely subcritical, critical, barely supercritical and very supercritical regimes or phases. For each of these phases $p$ has a specific definition. Or better said, these phases are defined by the difference in the probability $p$. The reason for there being two different subcritical and supercritical phases is because there is a rather noticeable difference if we change how $p$ gets close to $\frac{1}{n}$. The very subcritical and very supercritical phases are defined by the coarse parametrization

$$p = \frac{c}{n}.$$  

Where $c < 1$ for the very subcritical phase and $c > 1$ for the very supercritical phase, with $c$ a constant. The barely critical phases are defined by the fine parametrization

$$p = \frac{1}{n} + \lambda n^{-\frac{4}{3}} = \frac{1 + \epsilon}{n}.$$  

Where $\epsilon = \lambda n^{-\frac{4}{3}}$. The reason for using two notations is because sometimes expressing results in terms of $\epsilon$ works better, and sometimes expressing them in terms of $\lambda$ is best. For the very subcritical phase, $\lambda \to -\infty$, where for the very supercritical phase $\lambda \to \infty$. Note though that $\epsilon = \lambda n^{-\frac{4}{3}} \to 0$ for $n \to \infty$, or else $\frac{1 + \epsilon}{n}$ would be the same as $\frac{c}{n}$ for some $c \neq 1$ and the barely critical and very critical phases would be the same. For the critical window, we also have the fine parametrization but $\lambda$ is a real number.
We will denote the largest component as $C_1$ and the size (number of vertices in the component) by $L_1$. For the other components we use the same notations $C_i$ and $L_i$. It is clear that when $p$ is larger, $L_1$ will be larger as well. Here is an overview of how large the orders of $L_1$ and all $L_k$ for $k \neq 1$ fixed are across all phases:

<table>
<thead>
<tr>
<th>Phase</th>
<th>$L_1$</th>
<th>$L_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very subcritical</td>
<td>$\Theta(\ln n)$</td>
<td>$\approx L_1$</td>
</tr>
<tr>
<td>Barely subcritical</td>
<td>$\Theta(n^{\frac{2}{3}}\lambda^{-2}\ln \lambda)$</td>
<td>$\approx L_1$</td>
</tr>
<tr>
<td>Critical window</td>
<td>$\Theta(n^{\frac{2}{3}})$</td>
<td>$\approx L_1$</td>
</tr>
<tr>
<td>Barely supercritical</td>
<td>$2\lambda n^{\frac{3}{2}}$</td>
<td>$\Theta(n^{\frac{2}{3}}\lambda^{-2}\ln \lambda)$</td>
</tr>
<tr>
<td>Very supercritical</td>
<td>$yn$</td>
<td>$\Theta(\ln n)$</td>
</tr>
</tbody>
</table>

Figure 3.1.: An overview of the size of $L_1$ for the different phases of the phase transition in the random graph model.

The $y$ in the very supercritical phase is the positive real satisfying the equation $e^{-cy} = 1 - y$.

Proofs of all these results and more can be found in texts by Bollobás (2001)[6] and Janson et al (2000)[7]. The results that we are going to prove will be the size of the largest component for all phases and the uniqueness of the largest component in the supercritical phases. We will make one simplifying assumption for the barely supercritical phase though. Before we can prove these claims we will need to look at ways to analyze the random graph process. This will be done in the next few sections after which we will be able to prove our claims. First we will introduce the Breadth First Search algorithm and the Galton-Watson branching process, which will both be used for the branching models that we show after that. The first of those branching models represents our random graph and the other two are approximations of that model with modifications that make us able to do mathematics on them better. Then we will do an extensive analysis of those branching models, also comparing them to each other. We need this to finally be able to show the proofs of our claims for the different phases of the phase transition, starting with the subcritical phases, then the supercritical phases and finishing with the critical window.

However, before we jump right into the chapter, here are some more interesting results for the phase transition. Sudakov and Krivilevich [10] have found some easy proofs for interesting results for parts of the phase transition, which we will not prove here. These results include (for $\epsilon > 0$ a constant):

- Let $p = \frac{1-\epsilon}{n}$. Then whp all connected components of $\mathcal{G}(n,p)$ are of size at most $\frac{7}{\epsilon^2} \ln n$.
- Let $p = \frac{1+\epsilon}{n}$. Then whp there is a path of length at least $\frac{\epsilon^3 n}{5}$ in $\mathcal{G}(n,p)$.
- Let $p = \frac{1+\epsilon}{n}$. Then whp $\mathcal{G}(n,p)$ has a component of size at least $\frac{\epsilon n}{2}$. 

11
And at last, another interesting view at random graphs: Instead of constructing a random graph on $n$ vertices, Sudakov and Krivilevich took a graph in which every node has degree at least $n$ and studied a subgraph:

- Let $G$ be a finite graph with minimum degree at least $n$. Let $p = \frac{1+\epsilon}{n}$. Form a random subgraph $G_p$ of $G$ by including every edge of $G$ into $G_p$ with probability $p$. Then whp $G_p$ has a path of length at least $\frac{\epsilon^2 n}{5}$.

3.2. Breadth First Search algorithm and the Galton-Watson branching process

In the next section we will be giving three models that we will be analyzing for the rest of the chapter. For analyzing these models, we will need a way to analyze our random graph, which we will give in this section. The algorithm that we can use to search for components in a random graph is called the breadth first search (BFS) algorithm. We will also be using the Galton-Watson graph branching process, which will be explained a little further down.

3.2.1. Breadth First Search

We assume there is some order on our vertices. There are three sets of vertices called life vertices, neutral vertices and dead vertices. The live vertices are placed in a queue. At time $t = 0$ all vertices are neutral except for our starting vertex $v$. At step $t$ we first remove the vertex $w$ that is on top of the queue if it is nonempty. Then for all neutral vertices $w'$ we check if $w$ and $w'$ are neighbours. The vertex $w'$ is then placed at the bottom of the queue. Every time that the last live vertex is removed from the queue and there are only dead or neutral vertices, we have found a component $C(v)$ of our random graph and the algorithm ends. Figure 3.2 shows an example of a graph found by the BFS algorithm, with nodes numbered from 1 to 12 in the order of which they were placed in the queue.

![Figure 3.2.: An illustration of the order in which BFS finds nodes](image)

This algorithm only finds tree components, which is all we need as we only care about the size of a component and not the number of edges in it. Breadth first search first
looks for the nodes closest to the starting node. There is also an algorithm called Depth First Search (DFS), which first locates the nodes furthest away from the starting node. For example, in DFS the node with number 9 in Figure 3.2 would be node number 4. We will not give any further details of DFS as we will not use it for any proofs.

3.2.2. The Galton-Watson branching process

The Galton-Watson branching process is a stochastic process that we use to analyze the binomial and Poisson branching models, which we will discuss in the next section. Let \( Z \) be a distribution over the nonnegative integers. \( Z \) can be any distribution (the type of distribution does not matter for the construction of the process), but in the next section it will become clear that we are only looking at a Galton-Watson process with \( Z \) Poisson or binomially distributed. The first node in the Galton-Watson process is Eve. Eve has \( Z \) children. Each of her children also has \( Z \) children, independent of each other (note that this does not mean that all of Eve’s children have the same amount as her, but it means that all children use the same probability distribution). Those children have \( Z \) children as well, and so forth. Let \( T \) be the total number of offspring plus Eve. Depending on the probability distribution \( Z \), it is possible that \( T = \infty \).

Let \( Z_t \) for \( t \in \mathbb{N} \) be a sequence of i.i.d. variables, each with distribution \( Z \). We use the breadth first search algorithm to look at Eve and her offspring. This means that Eve is node 1 and she has \( Z_1 \) children, which are numbered 2, \ldots, \( Z_1 + 1 \), and those children’s children are numbered in the same manner, with node \( t \) having \( Z_t \) children. This corresponds to the Galton-Watson process, as \( Z_t \) are independent with distribution \( Z \).

We define the time \( t \) as the step in the process where the \( t \)th node has \( Z_t \) children and then dies, similar to the time when a vertex is removed from the queue in BFS. We call \( Y_t \) the number of alive vertices at time \( t \), meaning at \( t = 0 \) we have \( Y_t = 1 \) because at \( t = 0 \) Eve is alive. Then we have the recursive formula

\[
Y_t = Y_{t-1} + Z_t - 1.
\]

Because at time \( t \) the node \( t \) dies and \( Z_t \) new children are born. Now there are two possibilities:

1. \( Y_t = 0 \) for some \( t \). Let \( T \) be the smallest integer such that \( Y_T = 0 \). Then the Galton-Watson process stops after time \( T \) because all nodes have died. The total number of children, including Eve, is \( T \).

2. \( Y_t > 0 \) for all \( t \). Then the process goes on forever and \( T = \infty \), meaning the Eve family tree never ends.

Of course, whether the process will run infinitely or not depends on the probability distribution \( Z \).
3.3. Three branching models

As stated in the previous section, there are three models that we use to analyze random graphs. These models are in fact probability spaces, as is the random graph model. We want to analyze the graph branching model, which represents our random graph. The graph branching model is estimated by the binomial branching model, which is estimated by the Poisson branching model (we use the fact that the limit of a binomial distribution is a Poisson distribution). The reason for using the Poisson branching model is because it is the easiest to analyze. The graph branching process is comparable to the BFS algorithm and the binomial and Poisson process are comparable to a Galton-Watson process described in the previous section.

3.3.1. The graph branching process

The graph branching model is the model that we want to proof things for, as it represents our random graph \( G(n,p) \). The graph branching process start with a vertex \( v \) and finds the component \( C(v) \) of that vertex. However, it will not find all edges between vertices in \( C(v) \), but just a tree component. We have a probability space given by a sequence \( Z_1, \ldots, Z_n \), with \( Z_t \sim \text{Bin}(N_{t-1} - 1, p) \), \( N_0 = t - 1 \) and \( N_t = N_{t-1} - Z_t \). We also have the queue size \( Y_t \) at time \( t \) given by \( Y_0 = 1 \) and \( Y_t = Y_{t-1} + Z_{t-1} \). Let \( T^{gr} \) be the smallest \( t \) for which \( Y_t = 0 \). This \( T^{gr} \) is the size of the component \( C(v) \) as found by the BFS algorithm. In fact, the graph branching model mirrors the BFS algorithm until time \( T \), and then continues until time \( n \), which we call fictional continuation.

3.3.2. The binomial branching model

The binomial branching model is an approximation of the graph branching model. We use this model as a sort of middle ground, as we want to analyze the Poisson branching model, which is approximated by the binomial branching model and thus also the graph branching model. The binomial branching model is similar to the graph branching model, with a few changes. The changes are that \( Z_1, Z_2, \ldots \) now is an infinite sequence, and each \( Z_t \sim \text{Bin}(m, p) \), with \( m \) a positive integer. This means that instead of using a distribution on decreasingly fewer nodes \( n \), we keep using \( m \) nodes. Again we have the auxiliary variables \( Y_t \) given by \( Y_0 = 1 \) and \( Y_t = Y_{t-1} + Z_t - 1 \). However, in the binomial branching process there does not need to be a \( t \) for which \( Y_t = 0 \), in which case \( T^{bin}_{m,p} = \infty \). We still see \( Z_t \) as the number of nodes born at time \( t \) and \( Y_t \) as the queue size. Now \( T^{bin}_{m,p} \) is the total size of a Galton-Watson process with distribution \( Z = \text{Bin}(m, p) \).

3.3.3. The Poisson branching model

As said, the Poisson branching model is estimated by the binomial branching model. This is because the limit of a binomial \((n, p)\) distribution is a Poisson \((np)\) distribution. In this case we call the parameter of the Poisson distribution \( c, c \geq 0 \) and real. Like
in the binomial branching model we have an infinite sequence $Z_1, Z_2, \ldots, Z_t$ being the nodes born at time $t$, but this time they are Poisson distributed with mean $c$. We have the same recursion $Y_t = Y_{t-1} + Z_t - 1$, $Y_0 = 1$ for the queue size $Y_t$ and again $T^\infty$ is the smallest $t$ such that $Y_t = 0$, and $T^\infty = \infty$ if no such $t$ exists. Now $T^\infty$ is the size of a Galton-Watson process where $Z \sim \text{Po}(c)$.

### 3.4. analyzing the branching models

#### 3.4.1. The graph branching model

We take a look back at the BFS process. We take $Y_0 = 1$ and we have the recursion $Y_t = Y_{t-1} + Z_t - 1$. We also see that $N_0 = n - 1$ and $N_t = N_{t-1} - Z_t = n - t - Y_t$.

**Theorem 3.1.** In $G(n, p)$

$$\Pr[|C(v)| = t] \leq \Pr[\text{Bin}(n - 1, 1 - (1 - p)^t) = t - 1].$$

**Proof.** For each $t$, $Z_t$ is found by checking $N_{t-1}$ pairs of vertices for adjacency. As each pair is adjacent with probability $p$ we have

$$Z \sim \text{Bin}(N_{t-1}, p) \sim \text{Bin}(n - (t - 1) - Y_{t-1}, p).$$

Using this in $N_t = N_{t-1} - Z_t$ yields $N_t \sim N_{t-1} - \text{Bin}(N_{t-1}, p) \sim \text{Bin}(N_{t-1}, 1 - p)$. We want to use induction to prove

$$N_t \sim \text{Bin}(n - 1, (1 - p)^t).$$

First we see that $N_0 = n - 1 \sim \text{Bin}(n - 1, 0)$ satisfies our hypotheses. Now suppose $N_k \sim \text{Bin}(n - 1, (1 - p)^k)$ for all $k = 1, 2, \ldots, t - 1$. Using that $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(X, q)$ implies $Y \sim \text{Bin}(n, pq)$ a total of $t$ times, we get $N_t \sim \text{Bin}(\ldots \text{Bin}(n - 1, 1 - p) \ldots)) \sim \text{Bin}(n - 1, (1 - p)^t)$.

If $T^\infty = t$, the number of dead vertices at time $t$ is $t$ and the number of live vertices is $0$, so $N_t = n - t$. This gives us the inequality

$$\Pr[|C(v)| = t] \leq \Pr[\text{Bin}(n - 1, (1 - p)^t) = n - t].$$

Note that this is an inequality, as there can be other times where $N_t = n - t$ because of fictional continuation. Now we use the binomial distribution

$$\Pr[\text{Bin}(n - 1, (1 - p)^t) = n - t] = (n - t)(1-p)^t \left(\frac{n - 1}{n - t}\right) (t - 1)^{(1-p)^t}$$

$$= \Pr[\text{Bin}(n - 1, 1 - (1 - p)^t) = t - 1].$$

Which is exactly what we wanted to prove. \qed

For more precise bounds on $|C(v)|$ and an alternate analysis yielding the same result we refer to Van der Hofstad and Spencer (2006) [8]
3.4.2. Linking the models

**Theorem 3.2.** If \( c \leq 1 \), \( T_{c_{\infty}} \) is finite with probability one.

**Proof.** Suppose \( c < 1 \). for all \( t < T_{c_{\infty}} \) we have \( Y_t > 0 \) or \( \sum_{i=1}^{t} Z_i > t \). As \( Z_i \sim \text{Po}(c) \), \( \sum_{i=1}^{t} Z_i \sim \text{Po}(ct) \). Using the second inequality from B.1 with \( \epsilon = \frac{1}{c} - 1 \) we get

\[
\Pr[\text{Po}(ct) > t] = \Pr[\text{Po}(ct) > ct(1 + \epsilon)] < (e^{\frac{1}{c} - 1}c^{\frac{1}{c}})^{ct}.
\]

The term in the exponential goes to 0 as \( t \to \infty \) for \( c < 1 \), so \( \Pr[T_{c_{\infty}} > 0] \to 0 \), meaning \( T_{c_{\infty}} \) is infinite with probability 0, thus finite with probability 1.

Now suppose \( c \geq 1 \). We set \( \Pr[T < \infty] = z \). Suppose that in the branching process Eve has \( i \) children. The probability that the branching process is finite has probability \( z_i \), as all children of Eve must spawn a finite number of offspring. Thus we get

\[
z = \sum_{i=0}^{\infty} \Pr[Z_1 = i]z^i = \sum_{i=0}^{\infty} \frac{e^{-c}c^i}{i!}z^i = e^{-c} \sum_{i=0}^{\infty} \frac{c^i z^i}{i!} = e^{-c}e^{zc} = e^{(z-1)c} = e^{-yc}.
\]

Where we have set \( y = 1 - z \), which gives us the equation \( 1 - y = e^{-yc} \). Note that now \( y \) is the probability of \( T_{c_{\infty}} \) being infinity. If we have \( c = 1 \), then \( e^{-y} = 1 - y \). For \( y > 0 \) we know \( e^{-y} > 1 - y \), so this has only the solution \( y = 0 \), meaning the probability of \( T_{c_{\infty}} \) being finite is 1. \( \square \)

**Lemma 3.3.** Let \( c > 1 \). Let \( Z_1, Z_2, \ldots \) be independent and Poisson\((c)\) distributed. For \( a > 1 \) consider the process defined by \( Y_1 = a \) (meaning it is given that Eve has \( a \) children) and \( Y_t = Y_{t-1} + Z_t - 1 \) for \( t \geq 2 \). Then

\[
\lim_{a \to \infty} \sum_{t \geq 2} \Pr[Y_t \leq 0] = 0.
\]

**Proof.** First, note that \( Y_t = \sum_{i=1}^{t} Z_i - t = \sum_{i=2}^{t} Z_i - t + a = \tilde{Z}_t - t + a \), where \( \tilde{Z}_t = \sum_{i=2}^{t} Z_i \) is Poisson\((ct)\) distributed. Using Theorem A.1. we get

\[
\Pr[Y_t \leq 0] = \Pr[\tilde{Z}_t \leq t - a] = \Pr[\tilde{Z}_t \leq \frac{ct}{c} - \frac{a}{ct}dt] = \Pr[\tilde{Z}_t \leq ct(\frac{1}{c} - \frac{a}{ct})]
\]

\[
= \Pr[\tilde{Z}_t \leq ct(1 - (\frac{a}{ct} + \frac{1}{c}))] \leq e^{-\frac{1}{2}((\frac{a}{ct} + \frac{1}{c}))^2}
\]

\[
= e^{-\frac{1}{2}(\frac{a^2}{ct^2} + \frac{2a}{ct} + 1)} \leq e^{-\frac{a}{2} - \frac{t}{ct}}.
\]

Now the sum is

\[
S_{t,a} = \sum_{t \geq 2} \Pr[Y_t \leq 0] \leq \sum_{t \geq 2} e^{-\frac{a}{2} - \frac{t}{ct}}.
\]
We can move $e^{-\frac{a}{2}}$ in front of the sum, as it is independent of $t$. Because $\sum_{t \geq 2} e^{-\frac{a}{2}}$ converges and thus is a constant which we call $L$, we now have $S_{T,a} \leq Le^{-\frac{a}{2}} \to 0$ as $a$ goes to infinity.

**Theorem 3.4.** If $c > 1$, $T^p_c$ is infinite with probability $1 - y$, where $y(c)$ is the $y$ satisfying the equation $e^{-cy} = 1 - y$.

**Proof.** We use the proof of theorem 3.1, but now we look at $c > 1$. For this $c$ the function $f(y) = 1 - y - e^{-cy}$ has $f(0) = 0$, $f(1) = -e^{-c} < 0$ and $f'(0) = c - 1 > 0$. This means that there is a $y \in (0,1)$ with $f(y) = 0$. As $f$ is a convex function, there is exactly one such $y$. Now $\Pr[T^p_c < \infty] = 1$ or $\Pr[T^p_c < \infty] = 1 - y$. We are going to show that it is in fact the second of those.

Let $\epsilon > 0$ be a small constant. Because of lemma 3.2 there is an $A$ such that for all $Y_1 = a > A$ we have $\sum_{t \geq 2} \Pr[Y_t \leq 0] < \epsilon$. Since $\Pr[Y_1 > A]$ is nonzero (albeit very small) and the sum $\sum_{t \geq 2} \Pr[Y_t \leq 0] < \epsilon$, the probability that $Y_t > 0$ for all $t$ must be nonzero as well, meaning $T^p_c$ is infinite with probability greater than zero and thus with probability $1 - y$.

**Theorem 3.5.** For any positive real $c$ and any fixed integer $k$

$$\lim_{n \to \infty} \Pr[C(v) = k \text{ in } G(n, c/n)] = \Pr[T^p_c = k].$$

**Proof.** Let $\Gamma$ denote the set of $k$-tuples $\bar{z} = (z_1, \ldots, z_k)$ of nonnegative integers such that $y_t = y_{t-1} + z_t - 1$ with $y_0 = 1$ has $y_k = 0$ and $y_t > 0$ for $t < k$. Then we have the equalities

$$\Pr[T^g_c = k] = \sum_{\bar{z} \in \Gamma} \Pr[Z^g_{i} = z_i, 1 \leq i \leq k].$$

$$\Pr[T^p_c = k] = \sum_{\bar{z} \in \Gamma} \Pr[Z^p_{i} = z_i, 1 \leq i \leq k].$$

In words, we sum over all possible ways of reaching $y_k = 0$ without having $y_t = 0$ for a $t < k$. Now for one such $\bar{z}$

$$\Pr[Z^g_{i} = z_i, 1 \leq i \leq k] = \prod_{i=1}^{k} \Pr[Z^g_{i} = z_i].$$

Because $Z_i = n - O(1)$ ($k$ was fixed so it is $O(1)$) the binomial distribution $Z_i$ approaches the Poisson distribution

$$\lim_{n \to \infty} \Pr[Z^g_{i} = z_i] = \Pr[Z^p_{i} = z_i].$$

And as the product is of a fixed number of terms

$$\Pr[Z^g_{i} = z_i, 1 \leq i \leq k] = \Pr[Z^p_{i} = z_i, 1 \leq i \leq k].$$

Now we only need to note that $\Pr[C(v) = k] = \Pr[T^g_c = k]$ to complete the proof. 

17
**Theorem 3.6.** For any positive real \( c \) and any integer \( k \)

\[
\Pr[T^\text{po}_c = k] = \frac{e^{-ck}(ck)^{k-1}}{k!}.
\]

*Proof.* By theorem 3.5

\[
\Pr[T^\text{po}_c = k] = \lim_{n \to \infty} \Pr[|C(v)| = k].
\]

Where \( p = \frac{c}{n} \) and \( v \) is an arbitrary vertex of the graph. Since the graph has \( n \) vertices and the component has \( k \), one of which is \( v \), there are \( \binom{n-1}{k-1} \) choices for \( C(v) \). The probability that \( G(n, p) \) has more than \( k - 1 \) edges on the set \( S \) of \( C(v) \) is \( O(p^k) = O(n^{-k}) \). If the number of edges is exactly \( k - 1 \) then \( S \) is a tree component. A classical result from Cayley (1889)[9] tells us that the number of trees is \( k^{k-2} \). For each tree the probability of it occurring is \( p^{k-1}(1-p)^{(k-2)-(k-1)} \approx p^{k-1} = e^{k-1}n^{1-k} \) as \( (1-p)^{(k-2)-(k-1)} \approx 1 \). Thus the probability of having a connected component on \( S \) is \( e^{k-1}n^{1-k}k^{k-2} \). Because \( C(v) \) has exactly \( k \) vertices, there can be no edges between it and its complement. The probability of there being no such edges is \( (1-p)^{n-k} \) per edge, so \( (1-p)^{k(n-k)} \) total. Now as \( (1-p)^{-k} \approx 1 \)

\[
(1-p)^{k(n-k)} = ((1-p)^{n-k})^k = ((1 - \frac{e}{n})^n)^k = (e^{-c})^k = e^{-ck}.
\]

Now we use \( \binom{n-1}{k-1} = \frac{(n-1)!}{(k-1)!(n-k)!} \to \frac{n^{k-1}}{(k-1)!} \) so that the probability of finding a component of size \( k \) is

\[
\Pr[C(v) = k] = \binom{n-1}{k-1} e^{k-1}n^{1-k}k^{k-2}e^{-ck} \to \frac{e^{-ck}(ck)^{k-1}}{k!}.
\]

\[\square\]

**Theorem 3.7.** The probability of the Poisson branching process going until time at least \( u \) is bounded by the equation

\[
\Pr[T^\text{po}_c \geq u] < e^{-u(\alpha + o(1))}.
\]

*Proof.* We take a closer look at the equation that we just proved in theorem 3.6. Using Stirlings approximation \( n! \approx 2\pi k(\frac{n}{2\pi})^n \) we get

\[
\Pr[T^\text{po}_c = k] \approx \frac{1}{\sqrt{2\pi c}} k^{-\frac{3}{2}} (ce^{1-c})^k = \frac{1}{\sqrt{2\pi c}} k^{-\frac{3}{2}} e^{-ka}.
\]

(3.1)

Where \( \alpha = c - 1 - \ln c \). Now we can find the probability \( \Pr[T^\text{po}_c \geq u] = \sum_{k \geq u} \Pr[T^\text{po}_c = k] \).

Using that \( e^{-\alpha} \leq 1 \) for all \( c \neq 1 \), this goes to zero with exponential speed and thus every term is much smaller than the \( k = u \) term. This means we get the inequality

\[
\Pr[T^\text{po}_c \geq u] = \sum_{k \geq u} \frac{1}{\sqrt{2\pi c}} k^{-\frac{3}{2}} e^{-ka} < e^{-u(\alpha + o(1))}.
\]

\[\square\]
Now let us look at the Poisson branching process close to \( c = 1 \). First we look at \( c > 1 \). We parametrize \( c = 1 + \epsilon \). Remember the \( y \) from theorem 3.4. We want to solve
\[
1 - y - e^{-(1+\epsilon)y} = 0.
\]
Because of the implicit function theorem, there must be a solution \( y(\epsilon) \). We look at the series expansion of \( e^{-(1+\epsilon)y} \). Note that \( y = y(\epsilon) \) is a continuous function of \( \epsilon \) around 0. Since \( \epsilon \to 0 \) we get \( y \to 0 \). We can thus neglect any small terms of order greater than \( y^2 \) and \( \epsilon^2 \). Using this we get
\[
1 - y - e^{-(1+\epsilon)y} \approx 1 - y - (1 - (1 + \epsilon)y + \frac{1}{2}(1 + \epsilon)^2y^2) \approx cy - \frac{y^2}{2}.
\]
As this must be 0 we get
\[
y \approx 2\epsilon \text{ for } \epsilon \to 0^+.
\]
(3.2)

Now we first prove the equation for the probability of \( T_{1-\epsilon}^{po} > A\epsilon^{-2} \) being equal to \( A\epsilon^{-2} \). Why we choose \( A\epsilon^{-2} \) will become clear in the proof.

**Theorem 3.8.** For fixed \( A \) and \( \epsilon \to 0^+ \)
\[
\Pr[T_{1-\epsilon}^{po} \geq A\epsilon^{-2}] = \epsilon e^{-(1+o(1))\frac{A}{2}}.
\]

**Proof.** Suppose \( \epsilon \to 0^+ \). Using Taylor’s Theorem we have \( c e^{1-c} = (1 + \epsilon)(1 - \epsilon + \frac{\epsilon^2}{2} + O(\epsilon^3)) = 1 - \frac{\epsilon^2}{2} + O(\epsilon^3) \). Thus (3.1) gives us
\[
Pr[T_{1+\epsilon}^{po} = u] \approx \frac{1}{2\pi}u^{-\frac{3}{2}}.
\]

For \( u = o(\epsilon^{-2}) \). When \( u \) reaches order \( \epsilon^{-2} \), say \( u = A\epsilon^{-2} \) with \( A \) fixed we get
\[
Pr[T_{1+\epsilon}^{po} = A\epsilon^{-2}] \approx \frac{1}{2\pi}e^3A^{-\frac{3}{2}}e^{-\frac{A}{2}}.
\]

Now if we take \( A \to \infty \) we can put the smaller factors in the exponential term to get
\[
Pr[T_{1+\epsilon}^{po} = A\epsilon^{-2}] \approx e^3e^{-(1+o(1))\frac{A}{2}}.
\]

Now for \( c \) just a bit smaller than 1. We parametrize \( c = 1 - \epsilon \). Just like for \( 1 + \epsilon \) we have \( c e^{1-c} \approx 1 - \frac{\epsilon^2}{2} \). So when \( u = o(\epsilon^{-3}) \) we have
\[
Pr[T_{1+\epsilon}^{po} = A\epsilon^{-2}] \approx Pr[T_{1-\epsilon}^{po} = A\epsilon^{-2}] .
\]

And thus for \( A \to \infty \)
\[
Pr[T_{1-\epsilon}^{po} = A\epsilon^{-2}] = \epsilon^3e^{-(1+o(1))\frac{A}{2}}.
\]

So the Poisson branching process looks almost the same for \( 1+\epsilon \) and \( 1-\epsilon \). The difference is that for \( 1+\epsilon \) the process can be infinite, where it cannot be for \( 1-\epsilon \).

When \( \epsilon \to 0^+ \) and \( A \to \infty \)
\[
Pr[T_{1-\epsilon}^{po} \geq A\epsilon^{-2}] = \sum_{i \geq 1}^{r\epsilon^{-2}} Pr[T_{1-\epsilon}^{po} = A\epsilon^{-2} + i] < \sum_{i \geq 1}^{r\epsilon^{-2}} Pr[T_{1-\epsilon}^{po} = A\epsilon^{-2}] = r\epsilon^{-2} Pr[T_{1-\epsilon}^{po} = A\epsilon^{-2}] = \epsilon e^{-(1+o(1))\frac{A}{2}}.
\]
We only needed to sum over a constant times $\epsilon^{-2}$ terms (we put the constant $r$ in the $o(1)$ of in the exponential term) because the probabilities for terms of order greater than $\epsilon^{-2}$ are too small to notice. Thus we have proven the theorem.

\textbf{Theorem 3.9.} The probability of $T_{1+\epsilon}^{po}$ being infinite is

$$\Pr[T_{1+\epsilon} > A\epsilon^{-2}] \approx 2\epsilon = y.$$  

For $A \to \infty$ and $\epsilon \to 0^+$. 

\textit{Proof.} We use the proof of theorem 3.8. For the finite part of $T_{1+\epsilon}^{po}$ we get the same as for $T_{1-\epsilon}^{po}$

$$\Pr[\infty > T_{1+\epsilon}^{po} > A\epsilon^{-2}] < \epsilon e^{-(1+o(1))\frac{A}{\epsilon}}.$$  

(3.3)

Because of 3.2 we know that $\Pr[\infty > T_{1+\epsilon}^{po} > A\epsilon^{-2}] + \Pr[T_{1+\epsilon}^{po} > \infty] \approx 2\epsilon$. When $A \to \infty$ 3.3 is $o(\epsilon)$. so when $A \to \infty$ and $\epsilon \to 0^+$ we get the desired result.

\textbf{Theorem 3.10.} For any $u$

$$\Pr[T_{n-u,p}^{bin} \geq u] \geq \Pr[T_{n,p}^{gr} \geq u] \geq \Pr[T_{n-u,p}^{bin} \geq u].$$

\textit{Proof.} First we prove the second inequality. For this proof we use an alteration to the graph branching process. When the number of vertices in a component (the number of live plus dead vertices) reaches $u$, we stop the process. This has no influence on the probability of reaching at least $u$ vertices. In our alternated process, after we remove a vertex $w$ from the top of the queue, we only check $n-1-s$ vertices for adjacency with $w$. Note that this is always possible as the amount of neutral vertices at any time is greater than $n-u$. This alternate graph process dominates the binomial $n-1-s$ process so the probability of reaching $u$ vertices is greater or equal than the binomial process.

Now for the first inequality. Like in the above proof we introduce a (different) modification to the graph branching process. Now, after we move a vertex $w$ from the queue, instead of checking all available $n-1-s$ vertices for adjacency with $w$, we create $s$ fictional vertices to also check for adjacency with $w$. This way the the component we will get is of size $T_{n-1,p}^{bin}$. The actual component $C(v)$ will be a subset of this component, meaning the binomial $n-1-s$ process dominates the graph branching process.

\section{3.5. The subcritical regimes}

\subsection*{3.5.1. The very subcritical phase}

Now that we are done analyzing the models, we can finally start with the proofs of the claims that we made in the beginning of this chapter. First, let us look at the very subcritical phase. We have the coarse parametrization $p = \frac{c}{n}$ with $c < 1$ a constant.

\textbf{Theorem 3.11.} In the very subcritical phase, the largest component $L_1$ is of order $\ln n$. 

20
Proof. We are going to use the first inequality of theorem 3.10:
\[
\Pr[T_{n,p}^{gr} \geq u] \leq \Pr[T_{n-1,p}^{bin} \geq u].
\]
From this we get with the Poisson approximation
\[
\Pr[|C(v)| \geq u] \leq (1 + o(1)) \Pr[T_c \geq u] < e^{-u(\alpha + o(1))}.
\]
With the last inequality being Theorem 3.7, with \(\alpha = c - 1 - \ln c\). We can see that this drops exponentially in \(u\), as \(\alpha > 0\). Now we can choose a \(K\) large enough (to be more precise, for \(K\) with \(K(\alpha + o(1)) > 1.01\)) such that if we take \(u = K \ln n\) we get
\[
\Pr[|C(v)| \geq u] < n^{-1.01}. \quad \text{This must hold for all } n \text{ vertices } v, \text{ and thus for any vertex } v
\]
we have
\[
\Pr[|C(v)| \geq u] < n \cdot n^{-1.01} = 1/n^{0.01} \to 0.
\]
And this means that the largest component \(L_1 = O(\ln n)\) a.s.

3.5.2. The barely subcritical phase

Now we will look at the barely subcritical phase. Here we have \(p = \frac{1-\epsilon}{n}\) with \(\epsilon = \lambda n^{-\frac{1}{4}}\).

**Theorem 3.12.** In the barely subcritical phase, the largest component \(L_1\) is of order \(Kn^{\frac{3}{2}}\ln^{-2} \lambda\).

**Proof.** As in the very subcritical phase we use Theorem 3.10 and the Poisson approximation to get
\[
\Pr[|C(v)| \geq u] \leq (1 + o(1)) \Pr[T_{1-\epsilon} \geq u].
\]
This time however, we cannot use Theorem 3.7 as that only holds for \(T_c\). So we will have to use Theorem 3.8 to get
\[
\Pr[T_{1-\epsilon} > A\epsilon^{-2}] < e^{-(1+o(1))\frac{A}{2}\epsilon}.
\]
Note that for this equation to hold we need to have \(\epsilon \to 0^+\) and \(A \to \infty\). We take \(A = K \ln \lambda\) and parametrize \(u = A\epsilon^{-2} = K \ln \lambda \epsilon^{-2} = Kn^{\frac{3}{2}}\lambda^{-2} \ln \lambda\). Now for \(K\) large enough our bound gives us
\[
\Pr[T_{1-\epsilon} \geq u] < e^{-3.1 \ln \lambda} = \epsilon \lambda^{-3.1}.
\]
Now let \(I_v\) be an indicator random variable with \(I_v = 1\) if \(|C(v)| \geq u\) and 0 otherwise. Then \(X = \sum_v I_v\) is the number of vertices in components of size at least \(u\). We get
\[
E[X] = nE[I_v] \leq n\epsilon \lambda^{-3.1} = n^{\frac{3}{4}}\lambda^{-2.1}.
\]
Now let \(Y\) be the number of components in the random graph of size at least \(u\). because \(Y \leq u^{-1}X\) we get
\[
E[Y] \leq u^{-1}X = K^{-1}\lambda^{-0.1}.
\]
As \(\lambda\) tends to \(\infty\) we have \(Y = 0\) a.s. which means the largest component \(L_1 \leq u = Kn^{\frac{3}{2}}\lambda^{-2} \ln \lambda\). \(\Box\)
3.6. The supercritical regimes

As in the subcritical regime we can look at two phases in the supercritical regime: The very supercritical phase and the barely supercritical phase. Similarities between these phases are that in both there emerges a dominant component, and all other components are relatively small, meaning the dominant component is unique. We will begin by looking at the very supercritical phase.

3.6.1. The very supercritical phase

Remember that for the very supercritical phase we have \( p = \frac{c}{n} \) with \( c > 1 \) a constant. All along we have claimed that in the supercritical phases a giant component emerges, but we never gave a formal definition of the giant component, which we will do now. Let \( y \) (dependent on \( c \)) be the positive real solution of \( e^{-cy} = 1 - y \). Let \( \delta \) be an arbitrarily small constant and let \( K \) be an appropriately large constant. Let \( L^+ = (y + \delta)n \) and \( L^- = (y - \delta)n \). We call a component \( C(v) \) giant if \( L^- < |C(v)| < L^+ \), small if \( |C(v)| < S := K\ln n \) and awkward if it is neither small nor giant.

**Theorem 3.13.** The probability of having an awkward component in \( G(n, p) \) where \( p = \frac{c}{n} \) with \( c > 1 \) is \( o(n^{-20}) \).

**Proof.** We have \( n - 2\delta n - K\ln N \approx n \) choices for \( t = |C(v)| \) and \( n \) choices for \( v \), meaning it suffices to show that for any \( v \) and any \( t \) that \( \Pr[|C(v)| = t] = o(n^{-22}) \). From theorem 3.1 it suffices to show that \( \Pr[\text{Bin}(n - 1, 1 - (1 - p)^t) = t - 1] = o(n^{-22}) \). When \( t = o(n) \), \((1 - \frac{c}{n})^t = \prod_{i=1}^{t} (1 - \frac{c}{n}) = 1 - \frac{ct}{n} + o(n^{-2}) \). As \( c > 1 \) this means that

\[
\Pr[\text{Bin}(n - 1, 1 - (1 - p)^t) = t - 1] < \Pr[\text{Bin}(n - 1, 1 - (1 - p)^t) \leq t - 1] \quad (3.4)
\]

\[
\approx \Pr[\text{Bin}(n - 1, \frac{cn}{t}) \leq t - 1]. \quad (3.5)
\]

As \( n \) is large, this binomial distribution approximates the Poisson distribution with mean \((n - 1)\frac{ct}{t} \). Then using Theorem A.2 we get

\[
\Pr[\text{Bin}(n - 1, \frac{cn}{t}) \leq t - 1] \approx \Pr[\text{Po}((n - 1)\frac{cn}{t}) \leq t - 1]
\]

\[
\leq \Pr[\text{Po}((n - 1)\frac{cn}{t}) \leq \frac{1}{2}(n - 1)\frac{cn}{t}] \leq e^{-\frac{1}{8}(n-1)\frac{cn}{t}}.
\]

Where we took \( \epsilon = \frac{1}{2} \) in the Chernoff bound. Because \( t = o(n) \) this is exponentially small in \( n \), meaning we can make it \( o(n^{-22}) \). If \( t \approx xn \) with \( x \) fixed, \( 1 - (1 - \frac{c}{n})^t = 1 - (1 - \frac{c}{n})^{nx} \approx 1 - e^{-cx} \). Because \( y \) is the unique solution to \( 1 - e^{-cy} = y \), for \( x \neq y \) we know that \( 1 - e^{-cx} \neq x \). This means that the mean \((n - 1)(1 - e^{-cx}) \neq nx \approx t \), so the mean of the binomial is not near \( t \) thus the probability of it being equal to \( t \) is exponentially small in \( n \), again meaning we can make it \( o(n^{-22}) \). Note that the '20' in this theorem can be any arbitrarily large number if we change \( K \). \( \square \)
The probability of having an awkward component in Theorem 3.15.

We call a component dominant if $|C(v)| > S$, and there is no awkward component not small is the same as giant. Theorem 3.7 gives us

$$\Pr[T_{n-S}^{\bin} \geq S] \leq \alpha \leq \Pr[T_{n-1}^{\bin} \geq S].$$

Because of the Poisson approximation $\Pr[T_{n-S}^{\bin} \geq S] \rightarrow \Pr[T_c^{\text{po}} \geq S]$ and $\Pr[T_{n-1}^{\bin} \geq S] \rightarrow \Pr[T_c^{\text{po}} \geq S]$, so $\alpha$ gets sandwiched between the two terms and thus $\alpha \approx \Pr[T_c^{\text{po}} \geq S]$.

As $c$ is fixed and $S \rightarrow \infty$ we have $\Pr[T_c^{\text{po}} \geq S] \approx \Pr[T_c^{\text{po}} = \infty] = y$, meaning $\alpha \approx y$. Thus $C(v)$ is giant with probability $y$. Since this probability is the same for all $v$, the expected number of vertices in giant components (we have not yet proved that there can only be one) is $\approx yn$. We will now prove that the giant component is unique.

We take $n^{-2} << p_1 << n^{-1}$, for example $p_1 = n^{-\frac{3}{2}}$. Let $G_1 \sim G(n,p_1)$ and $G \sim G(n,p)$ on the same set of $n$ vertices. Let $G^+ = G_1 \cup G \sim G(n,p^+)$ where $p^+ = p_1 + p - p_1 p$. We call this sprinkling: the relatively few edges of $G_1$ are sprinkled on $G$ to create $G^+$.

Note that the amount of edges that are in both $G$ and $G_1$ is neglectable. Suppose that $G$ has more than one giant component. Let $V_1$ and $V_2$ be the vertex sets of two of those components. There are $\Omega(n^2)$ pairs of vertices $\{v_1, v_2\}$ with $v_1 \in V_1$ and $v_2 \in V_2$. Since $p_1 >> n^{-2}$ the probability of one of these edges being in $G_1$ is $1 - o(1)$. If we add this edge the component $V_1 \cup V_2$ has size at least $2(y - \delta)n$. We have $p_1 << n^{-1}$ so $p^+ = p_1 + p - p_1 p \approx p = \frac{\epsilon n}{2}$. The probability of $G^+$ having a component this big is $o(n^{-20})$ according to theorem 3.9, thus the giant component is unique.

3.6.2. The barely supercritical phase

In the barely supercritical phase we have $p = \frac{1+\epsilon}{n}$ with $\epsilon = \lambda n^{-\frac{1}{2}}$ for $\lambda \rightarrow \infty$. We add the assumption $\lambda >> \ln n$. Note that $\epsilon^{-2} = \lambda^{-1} n^{-\frac{1}{2}} << 2\lambda n^{-\frac{1}{2}} = 2\epsilon n$.

Again, let $\delta$ be an arbitrarily small constant and $K$ an arbitrarily large constant. We set $S = K\epsilon^2 \ln n$, $L^- = (1 - \delta)2\epsilon n$ and $L^+ = (1 + \delta)2\epsilon n$. We call a component $C(v)$ small if $|C(v)| < S$, dominant if $L^- < |C(v)| < L^+$ and awkward otherwise. The reason for calling the component dominant instead of giant is because we used giant for a component of order $n$, which will not exist in the barely supercritical phase.

Theorem 3.15. The probability of having an awkward component in $G(n,p)$ where $p = \frac{1+\epsilon}{n}$ is $o(n^{-20})$.

We will not prove this theorem as the proof is analog to that for the very supercritical phase, with just some small details being different for $t$ of order greater than $o(n)$.

Theorem 3.16. In the barely supercritical phase, there is exactly one dominant component.

Proof. We set $\alpha = \Pr[|C(v)| \geq S]$ and as there is no awkward component not small is the same as dominant. Theorem 3.7 gives us

$$\Pr[T_{n-S}^{\bin} \geq S] \leq \alpha \leq \Pr[T_{n-1}^{\bin} \geq S].$$
Now we use the Poisson approximation so that $T_{n-1}^{\binom{n}{2}} \approx T_{n+\epsilon}^\po$. Because $S > \epsilon^2$ we get from Theorem 3.9

$$\alpha \leq \Pr[T_{1+\epsilon}^\po \geq S] \approx \Pr[T_{1+\epsilon}^\po = \infty] \approx 2\epsilon.$$ 

If we replace $n-1$ by $n-S$ the mean becomes $(S-1)p \approx Sn^{-1}$ lower. Because of our assumption that $\lambda \gg \ln n$ we have $Sn^{-1} = K\lambda^{-2}n^{-\frac{1}{3}} = o(\epsilon)$. So $T_{n-S}^{\binom{n}{2}}$ is approximated by $T_{1+\epsilon+o(\epsilon)}$ and we have

$$\alpha \geq \Pr[T_{1+\epsilon+o(\epsilon)}^\po \geq S] \approx \Pr[T_{1+\epsilon+o(\epsilon)}^\po = \infty] \approx 2\epsilon.$$ 

And thus $\alpha \approx 2\epsilon$. Now the expected number of vertices in dominant components is $2en$. The sprinkling argument for the uniqueness of the dominant component is analog to that of the giant component. \qed

### 3.7. The critical window

For the critical window we have, like for the barely critical phases $p = \frac{1}{n} + \lambda n^{-\frac{4}{3}}$. The difference being that for the critical phases $\lambda \to \pm \infty$, where for the critical window $\lambda \in \mathbb{R}$.

**Theorem 3.17.** Let $X$ be the number of tree components. In the critical phase the expected number of tree components between size $an^2$ and $bn^2$ with $a$ and $b$ fixed is

$$\lim_{n \to \infty} \mathbb{E}[X] = \int_a^b e^{A(c)}c^{-\frac{5}{2}}(2\pi)^{-\frac{1}{2}} dc. $$

**Proof.** We will begin by giving a formula for the number of tree components.

Fix $c > 0$ and let $X$ be the number of tree components of size $cn^{-\frac{2}{3}} = k$ in the random graph. Then

$$\mathbb{E}[X] = \left(\binom{n}{k}\right) k^{-2 - p} k^{-1}(1 - p)^{k(n-k) + \binom{k}{2} - (k-1)}. $$

This formula might not look obvious at first sight, so it requires some explaining. The $\binom{n}{k}$ is of course because there are $n$ vertices of which $k$ need to be in our component. The $p^{k-1}$ is also clear as a tree component of size $k$ has $k-1$ edges. Now for the power of $1 - p$. The $n(n-k)$ is because there cannot be an edge between the $k$ vertices in the component and the $n - k$ other vertices, and finally the $\binom{k}{2} - (k-1)$ is because there are $\binom{k}{2}$ possible edges in the component of which $k-1$ have to be in our graph and thus $\binom{k}{2} - (k-1)$ should not.

Now we want to simplify this formula. A trick often used here is to write $x$ as $e^{\ln x}$. Another thing we will do multiple times is the approximation $\ln(1-x) = x - \frac{x^2}{2} + O(x^3)$ for small $x$. We will work on this formula bit by bit, so first we’re going to work on $\binom{n}{k}$. We do this using Stirling’s approximation $n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{(n)(n-1)\ldots(n-k+1)}{k!} \approx \frac{n^k k^{-1}}{k!} \prod_{i=1}^{k-1} \left(1 + \frac{i}{n}\right) \approx \frac{n^k e^k}{k^k \sqrt{2\pi k}} \prod_{i=1}^{k-1} \left(1 + \frac{i}{n}\right).$$
And for \( i = 1, 2, \ldots, k - 1 \)
\[
-\ln(1 - \frac{i}{n}) = \frac{i}{n} + \frac{i^2}{2n^2} + O\left(\frac{i^3}{n^3}\right).
\]
Now we use the sums
\[
\sum_{i=1}^{k-1} \frac{i}{n} = \frac{k^2}{2n} + o(1) \quad \text{and} \quad \sum_{i=1}^{k-1} \frac{i^2}{2n^2} = \frac{k^3}{6n^2} + o(1)
\]
to get
\[
\sum_{i=1}^{k-1} -\ln(1 - \frac{i}{n}) = \frac{k^2}{2n} + \frac{k^3}{6n^2} + o(1) = \frac{k^2}{2n} + \frac{c^3}{6} + o(1).
\]
And putting all this together we have
\[
\binom{n}{k} = \frac{n^k e^k}{k^k \sqrt{2\pi k}} \prod_{i=1}^{k-1} \left(1 + \frac{i}{n}\right) = \frac{n^k e^k}{k^k \sqrt{2\pi k}} e^{-\frac{k^2}{2n} - \frac{c^3}{6} + o(1)}.
\]
Now for \( p^{k-1} = n^{1-k}(1 + \lambda n^{-\frac{1}{2}})^{k-1} \). With the same approximation of \( \ln \) as before and \( c = kn^{-\frac{3}{2}} \) we get
\[
(k - 1) \ln(1 + \lambda n^{-\frac{1}{2}}) = k\lambda n^{-\frac{1}{2}} - \frac{c\lambda^2}{2} + o(1).
\]
And thus
\[
p^{k-1} = n^{1-k} e^{(k-1)\ln(1+\lambda n^{-\frac{1}{2}})} = n^{1-k} e^{k\lambda n^{-\frac{1}{2}} - \frac{c\lambda^2}{2} + o(1)}.
\]
And for the last part we have
\[
(1 - p)^{k(n-k)+\binom{k}{2}-(k-1)} = e^{(k(n-k)+\binom{k}{2}-(k-1))\ln(1-p)}.
\]
First we approximate the \( \ln \) term
\[
\ln(1-p) = -p + O(n^{-2}) = -\frac{1}{n} - \frac{\lambda}{n^{\frac{3}{2}}} + O(n^{-2}).
\]
Then the other term in the exponential
\[
k(n-k) + \binom{k}{2} - (k-1) = k(n-k) + \frac{k(k-1)}{2} - (k-1) = kn - \frac{k^2}{2} + O(n^3).
\]
And putting those two together gives us
\[
(k(n-k) + \binom{k}{2} - (k-1))\ln(1-p) = -k + \frac{k^2}{2n} - \frac{\lambda k}{n^{\frac{1}{2}}} + \frac{c\lambda^2}{2} + o(1).
\]
So that
\[
(1 - p)^{k(n-k)+\binom{k}{2}-(k-1)} = e^{-k + \frac{k^2}{2n} - \frac{\lambda k}{n^{\frac{1}{2}}} + \frac{c\lambda^2}{2} + o(1)}.
\]
Now we can put all our three parts together to get an equation for \( \mathbb{E}(X) \)
\[
\mathbb{E}(X) \approx -\frac{n^k k^{-2}}{k^k n^{k-1} \sqrt{2\pi k}} e^A = nk^{-\frac{5}{2}} (2\pi)^{-\frac{1}{2}} e^A.
\]
Where
\[ A = k - \frac{k^2}{2n} - \frac{c^3}{6} + \frac{k\lambda}{n^{\frac{3}{2}}} - \frac{c\lambda^2}{2} - k + \frac{k^2}{2n} - \frac{\lambda c^2}{2} + \frac{\lambda c^2}{2} - \frac{\lambda c^2}{2} = \frac{(\lambda - c)^3 - \lambda^3}{6}. \]

Meaning \( A \) is dependent only on \( c \) and \( \lambda \). As we usually take a fixed \( \lambda \), we can write \( A = A(c) \). If we now write \( k \) in terms of \( n \) the expected number of tree components becomes
\[ \mathbb{E}[X] = n^{-\frac{3}{2}}e^{A(c)}c^{-\frac{5}{2}}(2\pi)^{-\frac{1}{2}}. \]

Now for any \( k \) this limit goes to 0. However, we can sum \( k \) between \( cn^{\frac{3}{2}} \) and \( (c + dc)n^{\frac{3}{2}} \), meaning we have to multiply by \( n^{\frac{3}{2}}dc \) because we sum over that amount of \( k \). Now if we take the limit of \( n \to \infty \) we get an integral. Let \( X \) be the number of tree components of size between \( an^{\frac{3}{2}} \) and \( bn^{\frac{3}{2}} \). Then we get the integral
\[ \lim_{n \to \infty} \mathbb{E}[X] = \int_a^b e^{A(c)}c^{-\frac{5}{2}}(2\pi)^{-\frac{1}{2}} dc. \]

However, this is only the amount of tree components. For non-tree components we will not give any proofs, these can be found in the works of Wright (1997) [11]. we will however give the integral for the expectation. For fixed \( a, b, \lambda, \ell \) the number \( X^{(\ell)} \) of components of size between \( an^{\frac{3}{2}} \) and \( bn^{\frac{3}{2}} \) with \( \ell - 1 \) more edges than vertices satisfies
\[ \lim_{n \to \infty} \mathbb{E}[X] = \int_a^b e^{A(c)}c^{-\frac{5}{2}}(2\pi)^{-\frac{1}{2}}(c_\ell c^{\frac{3\ell}{2}}) dc. \]

Where \( c_\ell \) is given by a special recurrence. \( c_0 = 1, c_1 = \sqrt{\frac{3}{8}} \) and \( c_\ell = \ell^{\ell(1+o(1))} \) asymptotically in \( \ell \). Then if \( X^* \) is the total number of components of size between \( an^{\frac{3}{2}} \) and \( bn^{\frac{3}{2}} \) and \( g(c) = \sum_{\ell=0}^\infty c_\ell c^{\frac{3\ell}{2}} \) we get
\[ \lim_{n \to \infty} \mathbb{E}[X^*] = \int_a^b e^{A(c)}c^{-\frac{5}{2}}(2\pi)^{-\frac{1}{2}} g(c) dc. \]
4. Other random graph models and continuity

A paper by Achlioptas, D’Souza and Spencer (2009) [12] has generated a lot of interest in explosive percolation. Explosive percolation is when a large connected component emerges in a number of steps much smaller than the size of the system, like in Achlioptas-processes, which the random graph model $G(n,m)$ is a subset of. Since then, it has been proven that explosive percolation is continuous. One proof of the continuity is by Da Costa, Dorogovtsev, Goltsev, and Mendes (2010) [17]. However, to quote Riordan and Warnke (2011) [14]: Their argument applies only to one specific rule [known by their initials (dCDGM)] and is a (sophisticated) heuristic, not a proof. In particular, they start by assuming a continuous transition, and show that this is self-consistent. Reason enough for Riordan and Warnke to construct their own proof, which we will show in this section.

4.1. Some other random graph models

In figure 4.1 is an overview of the phase transition in each of the random graph models we will be naming here.

For each model we will either give a short description or a reference to find more about the model. Some models are Achlioptas or Achlioptas-like processes, which we will be discussing in the next section. The models are ordered like in the legend of the graph: from earliest phase transition to latest.

- Dubins, see Bollobás, Janson & Riordan (2007) [15]. Named after the three authors, this model falls in the BJR family.
- Power-Law, see Bollobás, Janson & Riordan (2007) [15]. Like Dubins, PL is in the BJR family.
- Erdős-Renyi - $G(n,p)$ or $G(n,m)$. ER also falls in the BJR family.
- Bohman-Frieze - we take $e_1$ and $e_2$ and choose $e_1$ if it connects two vertices with degree 0, and $e_2$ otherwise. This is an Achlioptas process.
- Adjacent edge rule - D’Souza & Mitzenmacher (2010) [16]. AE is an Achlioptas-like process.
- Triangle rule - D’Souza & Mitzenmacher (2010) [16]. TR is an Achlioptas-like process.
• Sum rule - take $e_1$ and $e_2$ and choose the edge that creates the smallest component. SR is an Achlioptas process.

• Product rule - take $e_1$ and $e_2$ and choose the edge that has the smallest product of the size of its vertices' components. Like SR, PR is also an Achlioptas process.

• dCDGM - Named after its inventors Da Costa, Dorogovtsev, Goltsev and Mendes (2010) [17], the dCDGM process is an Achlioptas-like process.

All of these processes have a continuous phase transition, even though figure 3.3 might not make that obvious for the 5 last models. However, there is one model that has a discontinuous phase transition. For this model we add one edge at every step $t$. The edge that gets added to the graph is the one that creates the smallest component. It is clear that in this way, all components will stay small until the very end of the process, at which all components suddenly become connected to each other, meaning the size of the largest component grows extremely fast in the last few steps. It turns out that the phase transition is no longer continuous for the following rule: Whenever $\ell \to \infty$ as $n \to \infty$, the rule 'pick $\ell$ vertices at random and join the two smallest distinct components selected’. Proof of this can be found in the text of Riodan & Warnke (2012) [13].

4.2. Continuity of the phase transition

For this section we will be looking at the $\mathcal{G}(n,m)$ model instead of the $\mathcal{G}(n,p)$ model, as the comparison between this model and other random graph models is clearer. All proofs and results in this section are from Riordan & Warnke (2012) [13].
Now for the continuity of the phase transition. By this we mean that the growth of the largest component $L_1$ of a random graph is continuous. Before we can give the formal definition, we need to define some other things to help us, starting with the $\ell$-vertex rule. For each $n$, let $(v_1, v_2, \ldots)$ be an i.i.d. sequence, where each $v_m$ is a sequence $(v_{m,1}, \ldots, v_{m,\ell})$ of $\ell$ vertices from the set of $n$ vertices. An $\ell$-vertex rule is a random sequence $(G(m))_{m \geq 0}$ on the set of $n$ vertices satisfying

1. $G(0)$ is the empty graph on $n$ vertices.
2. For every $m \geq 1$, $G(m)$ is formed by adding a set $E_m$ of edges to $G(m - 1)$, with all edges in $E_m$ being between vertices in $v_m$. Note that $E_m$ is allowed to be empty.
3. If all $\ell$ vertices in $v_m$ are in distinct components of $G(m - 1)$ then $E_m$ cannot be empty.

The set $E_m$ can be chosen in multiple ways. For the $G(n,m)$ model $\ell = 2$ and $E_m$ is always one edge. Other ways to choose $E_m$ lead to other random graph models. In the previous section we also mentioned Achlioptas processes. An Achlioptas process is an $\ell$-vertex rule where there are always exactly two possible edges $e_1$ and $e_2$ to choose from and at every step exactly one edge will be added to the graph.

Now that we have $\ell$-vertex rules, we can give a formal definition of continuity of the phase transition and give a proof that is in fact continuous. Because we will prove the continuity for all $\ell$-vertex rules, this means all random graph models made from $\ell$-vertex rules are continuous.

**Definition 4.1.** Let $R$ be an $\ell$-vertex rule with $\ell \geq 2$. For each $n$, let $(G_R(n,m))_{m \geq 0}$ be the random sequence of graphs on the set of $n$ vertices associated to $R$. Let $h_L$ and $h_m$ be functions that are both $o(n)$ and let $\delta > 0$ be any constant. Then the growth of the largest component $L_1$ is said to be continuous if the probability that there exist $m_1$ and $m_2$ with $L_1(G(m_1)) \leq h_L(n)$, $L_1(G(m_2)) \geq \delta n$ and $m_2 \leq m_1 + h_m(n)$ tends to 0 as $n \to \infty$.

So we actually defined the growth of $L_1$ being continuous as it not being discontinuous. Now we have the following theorem.

**Theorem 4.2.** The growth of the largest component $L_1$ is continuous for every $\ell$-vertex rule.

To prove this theorem we first need two lemmas, which we will not prove. Proof of the lemmas can be found in the works of Riordan & Warnke (2012).

**Lemma 4.3.** Let $N_{\geq k}$ be the number of vertices of a graph $G$ in components of size at least $k$. Given $0 < \alpha \leq 1$, let $C(\alpha)$ denote the event that for all $0 \leq m \leq n^2$ and $1 \leq k \leq \frac{\alpha}{16} \frac{n}{\ln n}$, the following holds: $N_{\geq k}(m) \geq \alpha n$ implies $L_1(m + \Delta) > \frac{\alpha}{\pi} n$ for $\Delta = \left\lceil \frac{4}{\alpha} \frac{n^2}{k} \right\rceil$. Then $\Pr[C(\alpha)] \geq 1 - \frac{1}{n}$.

In words, this lemma says that if there are at least $\alpha n$ vertices in components of size at least $k$, after time $\Delta$ the largest component of the graph will be of size at least $\frac{\alpha}{\pi} n$ with probability at least $1 - \frac{1}{n}$.
Lemma 4.4. Fix $0 < \alpha \leq 1$, $D > 0$ and $B \in \mathbb{N}$ with $B \geq 2$. Define $M^B_k(m) = N_{\geq k} - N_{\geq Bk}(m)$. Let $\mathcal{L}(\alpha, B, D)$ denote the event that for all $0 \leq m \leq n^2$ and $0 \leq k \leq \min\left\{\frac{\alpha e^{-2tBD}}{2B}, \frac{n}{2B}\right\}$ the following holds: $M^B_k(m) \geq \alpha n$ implies $M^B_k(m+\Delta) > \frac{\alpha n}{2B} e^{-2tBD}$ for every $0 \leq \Delta \leq \frac{Dn}{k}$. Then $\Pr[\mathcal{L}(\alpha, B, D)] \geq 1 - \frac{1}{n}$.

This lemma says that if there are at least $\alpha n$ vertices in components of size between $k$ and $Bk$, then after time $\Delta$ with probability at least $1 - \frac{1}{n}$ there will still be $\alpha n e^{-2tBD}/2B$ vertices in components of size between $k$ and $Bk$.

Proof of theorem 4.2. Let $X = X_n(\delta, h_L, h_m)$ denote the event that there exist $m_1$ and $m_2$ satisfying $L_1(g(m_1)) \leq h_L(n)$, $L_1(G(m_2)) \geq \delta n$, and $m_1 \leq m_2 + h_m(n)$. Now we need to show that the probability of $X$ tends to 0 as $n \to \infty$. We shall define a good event $H = H_n(\delta)$ such that $\Pr[H] \to 1$ as $n \to \infty$ and show that there is some $n_0$ such that for $n \geq n_0$ when $H$ holds, $X$ does not.

We set $\alpha = \frac{\delta}{4}$, $A = \frac{\alpha e^{-2tBD}}{2B}$, $D = 1$, $B = \left\lceil \frac{2AC}{\delta} \right\rceil$, $\beta = \frac{\alpha e^{-2tBD}}{2B}$, and $K = B^1[\frac{1}{2}]$.

Now we use lemmas 4.3 and 4.4. Let $H$ be the event that $C(1), C(\alpha)$ and $\mathcal{L}(\alpha, B, D)$ all hold. Our lemmas tell us that $\Pr[H] \geq (1 - \frac{1}{n})^3 = 1 - o(1)$. Because of our lemmas and the definition of $H$, if $n$ is large enough then for all $m \leq 5n$ and $k \leq K$ the following two statements hold:

- (1) $N_{\geq k} \geq \alpha n$ implies (2) $L_1(m + [\frac{4n}{k}]) \geq \frac{\alpha n}{2B}$.
- (3) $M^B_k(m) \geq \alpha n$ implies (4) $M^B_k(m') \geq \beta n$ for all $m' \leq m + \frac{n}{k}$.

Suppose that $H$ holds and that $m^- = \max\{m : L_1(G(m)) \leq h_L(n)\}$ and $m^+ = \min\{m : L_1(G(m)) \geq \delta n\}$ differ by at most $h_m(n)$. It suffices to show that if $n$ is large enough this gives a contradiction.

Because $N_1(0) = n$ and $C(1)$ holds, we have $L_1(4n) \geq \frac{n}{2B}$. For $n$ large enough it follows that $m^- \leq 4n$, so $m^+ \leq 5n$. For $k \leq \frac{1}{B}$ set $m_k = m^+ - \frac{\delta n}{4}$. Note that this is positive, as required. If we go from $G(m^k)$ to $G(m^+)$, at most $\left(\frac{\delta}{2}\right)(m^+ - m_k) < \frac{\ell^2}{2}(m^+ - m_k)$ edges are added. This means that the components of $G(m_k)$ of size at most $k$ contribute at most $k\frac{\ell^2}{2}(m^+ - m_k) \leq \frac{\delta n}{4}$ vertices to any one component of $G(m^+)$. It follows that $N_{\geq k}(m_k) \geq L_1(G(m^+)) - \frac{\delta n}{2} \geq \frac{\delta n}{2}$. Now suppose that $N_{\geq Bk}(m_k) \geq \frac{\delta n}{4}$. Then (1) holds at step $m_k$ with $Bk$ in place of $k$, so we get from (2) that at time $m^+ = m_k + \left[\frac{4n}{Bk}\right] \leq m_k + \frac{4n}{Bk} = m^+ - \frac{\delta n}{4} = m^+ - O(n)$ we have $L_1(G(m^+)) \geq \frac{\alpha n}{2B} > h_L(n)$ for large $n$. Since $m^+ - m^- = h_m(n) = o(n)$, for large enough $n$ we have $m^* < m^-$. But $L_1(G(m^+)) < h_L(n) < L_1(G(m^*))$ means $m^* > m^-$, meaning this is a contradiction.

It follows that $M^B_k(m_k) = N_{\geq k}(m_k) - N_{\geq Bk}(m_k) \geq \alpha n$. Because (3) implies (4) this gives $M^B_k(m^+) \geq \beta n$. If we do this for $k = 1, B, B^2, \ldots, B^\left\lceil\frac{1}{\beta}\right\rceil$, meaning a total of $\left\lceil\frac{1}{\beta}\right\rceil$ times, the total amount of vertices in the graph will be $\beta n \left\lceil\frac{1}{\beta}\right\rceil > n$. This is of course a contradiction, as we only have $n$ vertices in our graph.
5. Popular summary

A graph is a number $n$ of vertices with an edge set $E$ on those vertices. In graphs two vertices are connected if there is an edge between those vertices. By walking over edges from one vertex to another we get paths, and if a path has the same endpoint as starting point and never crosses the same vertex twice, it is called a cycle. If we take the collections of all edges that are connected to each other, we have a component of a graph. If we color every vertex in a graph, and give two connected vertices different colors we get a colored graph. If the smallest number of colors to be able to color a graph is $k$, the graph is called $k$-colorable.

![Figure 5.1.: All possible graphs on 3 edges](image)

There are multiple ways to construct random graphs. The first is to have $n$ vertices and for every pair on those $n$ vertices we can draw an edge. Each possible edge is then included in the graph with probability $p$, meaning the graph we get is a random graph on $n$ vertices. Another way is to look at all possible graphs at $n$ vertices with $m$ edges and randomly take one of them. The third method is to start with $n$ vertices and no edges, and then one by one add a random edge to the graph. All three methods result in random graphs that behave similar. If we look at figure 5.1, the first method has probability $p^3$ to choose the first graph, $p^2(1-p)$ for all three graphs with two edges, $p(1-p)^2$ for the graphs with one edge and $(1-p)^3$ for the empty graph. The second method has only one possibility if $m = 0$ and also if $m = 3$, and if $m = 1$ it has three possible outcomes, just like for $m = 2$. However, these graphs are isomorphic so we consider them to be the same graph. The third method starts with the empty graph and then, depending on the maximum number of edges we add, transitions to a graph...
with one edge, then one with two edges and finally the complete graph with all three edges. We gave an example with only three vertices, but when studying random graphs we usually let \( n \) tend to infinity.

One application of random graphs is the probabilistic method. This method is used to prove the existence of a kind of mathematical object, without constructing it. A classical example of this is the theorem by Erdős that says that for every number \( k \) there exists a graph in which the smallest cycle has length \( k \), but the graph is also \( k \)-colorable. This is counterintuitive as no small cycles means that locally the graph is 2-colorable.

When we look at the model in which we add \( m \) edges, something interesting happens. If \( m \) is slightly smaller than \( \frac{n^2}{2} \), the largest components of the random graph are of size a constant times \( \ln n \). Keep in mind \( n \to \infty \), so a constant times \( \ln n \) is always smaller than every positive power of \( n \). Then if we slowly let \( p \) grow until it is just slightly smaller than \( \frac{n^2}{2} \), there is one giant component with size a constant times \( n \), while all other components are still of order \( \ln n \). The emergence of this giant component has been intensively studied by mathematicians for the past half century ever since Erdős found its existence in 1960.

There are some other random graph models. An example is the Bohman-Frieze model, which starts with an empty graph and adds edges one by one. However, the edge added is not completely random. To find which edge we add, we take two random edges. If the first edge is between two vertices that are not connected to any other vertices, we take that edge. If not, we take the other edge and add it to the graph. This is just one of multiple models which changes the way the random graph is constructed. As the model is different, the number of edges at which the phase transition occurs is different as well. For the Bohman-Frieze model the phase transition occurs when there are around \( \frac{3n}{5} \) edges added to the graph.

Lastly, the phase transition appears to be continuous. This means that the growth of the largest component is not explosive. This is true for all but one random graph model. The only random graph model that has an explosive growth of the largest component is this particular one: Start with an empty graph, and add edges one by one. The edge that gets added, is the edge that connects the two smallest components in the graph (remember that if two components are equally small it does not matter which edge we take as the graphs are isomorphic).
Bibliography


A. Probability distributions and Chernoff bounds

First we give the definitions of the binomial and Poisson distributions. The binomial distribution is the discrete probability distribution of the number of successes in a sequence on \( n \) independent experiments, all of which yield 1 with probability \( p \) and 0 with probability \( 1 - p \). The probability of getting exactly \( k \) successes is

\[
\Pr[\text{Bin}(n,p) = k] = \binom{n}{k} p^k (1 - p)^{1-k}.
\]

This is the probability mass function. The mean of the binomial distribution is \( np \) and its variance is \( np(1 - p) \).

The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and if these events occur with a known average rate and independently of the time since the last event. The only parameter of the Poisson distribution is \( \lambda \) and its probability mass function is

\[
\Pr[\text{Po}(\lambda) = k] = e^{-\lambda} \frac{\lambda^k}{k!}.
\]

The mean of the Poisson distribution is \( \lambda \), and so is its variance.

The reason for using these two distributions, is because of the Poisson approximation. In short, the Poisson approximation says that the limit of a binomial distribution is a Poisson distribution. Formally:

**Theorem A.1.** If \( n \to \infty \), \( p \to 0 \) such that \( np \to \lambda \) then

\[
\frac{n!}{k!(n-k)!} p^k (1 - p)^{n-k} \to e^{-\lambda} \frac{\lambda^k}{k!}.
\]

**Proof.** Using Stirling’s approximation \( n! \to \sqrt{2\pi n \left(\frac{n}{e}\right)^n} \) we get

\[
\frac{n!}{k!(n-k)!} p^k (1 - p)^{n-k} \to \frac{\sqrt{2\pi n \left(\frac{n}{e}\right)^n}}{\sqrt{2\pi (n-k) \left(\frac{n-k}{e}\right)^{n-k} k!}} \to \frac{n^n p^k (1 - p)^{n-k}}{(n-k)^{n-k} e^{k k!}}.
\]

Now using \( np \to \lambda \), \( (1 + \frac{x}{n})^n \to e^x \) and \( (1 - \frac{x}{n})^k \to 1 \) we get

\[
\frac{n^n p^k (1 - p)^{n-k}}{(n-k)^{n-k} e^{k k!}} \to \frac{n^n \left(\frac{\lambda}{n}\right)^k (1 - \frac{\lambda}{n})^{n-k}}{(1 - \frac{\lambda}{n})^{n-k} e^{k k!}} \to \frac{\lambda^k (1 - \frac{\lambda}{n})^n}{(1 - \frac{\lambda}{n})^{n-k} e^{k k!}} \to \frac{\lambda^k e^{-\lambda}}{e^{-\lambda} e^{k k!}} = e^{-\lambda} \frac{\lambda^k}{k!}.
\]
Another way of using the Poisson approximation is by saying the Bin\((n, p)\) distribution can approximate the Po\((cp)\) distribution as close as we want. If \(n \to \infty\) then for any \(u\),

\[
\Pr[\text{Bin}(n, p) \geq u] = (1 + o(1)) \Pr[\text{Po}(cp) \geq u]
\]

The reason we use the Poisson distribution is because we have bounds for it. The following are called Chernoff bounds

**Theorem A.2.** Let \(P\) be \(\text{Poisson}(\mu)\) distributed. For \(\epsilon > 0\)

\[
\Pr[P \leq \mu(1 - \epsilon)] \leq e^{-\frac{1}{2}(1 - \epsilon)\mu}.
\]

and

\[
\Pr[P \geq \mu(1 + \epsilon)] \leq (e^{\epsilon}(1 + \epsilon)^{-(1 + \epsilon)})^{\mu}.
\]

There are many more Chernoff bounds, but we only need these two.