The phase transition behavior of random graphs

Manuele Leonelli

September 6, 2011

Abstract

In the last few years, because of the increasing importance of the study of complex networks such as the World Wide Web or social networks, the study of random graphs has experienced a great interest from the scientific community. Indeed, they are a very simple model which can satisfactorily represent the application considered, usually quite complex. In this dissertation we focus on a fascinating property of random graphs: a phase-transition behavior for the size of the largest component. Many different proofs of this behavior have been proposed, using quite different approaches. In the last few years a new approach has been proposed, which is based on quite basic probability concepts and it is much easier than the other proofs developed. In particular, the main point of this new approach is to compare, in a way that will be explained in the dissertation, components of a random graph to branching processes and to exploit the characteristics of this stochastic process. For this reason, the whole second chapter is devoted to the study of branching processes. Moreover, the analysis of the size of the components requires some concepts of probability theory, such as large deviation theory and stochastic domination, which are not usually studied by Master students. These introductory topics are introduced in the first chapter, together with a brief review of the models of random graphs and a short summary of graph theory's terms. Thus, after the first two "introductory" chapters, in the third one is developed the study of the phase transition behavior of random graphs. The main aim of the work is to present the theorems and their proofs about this topic in a comprehensive and accessible way, giving summaries and detailed explanations of the main steps of the work.

Contents

Chapter 1

Preliminaries

In this first chapter we will give some basic results about the main areas of study of this project, that will be exploited in the rest of the work. In the first section we will review some concepts of graph theory, in the second one we will introduce the notion of random graph, while in the last we will consider some probability concepts.

1.1 Graph theory

This section is based on the bibliographic references [1] and [2] and in particular on the first two chapters of each of them. We will basically introduce all the terms of graph theory that will be used in the following. Obviously, we are going to start by explaining what a graph is.

A **graph** is a pair $G = (V, E)$ of sets satisfying $E \subseteq [V]^2$: thus, the elements of E are 2-element subsets of V. The elements of V are the **vertices** (or **nodes**) of the graph G, while the elements of E are its **edges**. The number of vertices of a graph G is its **size**. The usual way to picture a graph is by drawing a shape (it can be, for example, a dot or an ellipse as in Figure 1.1) for each vertex and by joining together two of these shapes with a line if the corresponding two vertices form an edge. In Figure 1.1 we can see an example of a graph with vertex set $V = \{1, 2, 3, 4, 5, 6, 7\}$ and edge set $E = \{\{1, 3\}, \{2, 3\}, \{1, 5\}, \{2, 5\}, \{4, 6\}, \{5, 6\}, \{3, 7\}, \{4, 7\}\}$.

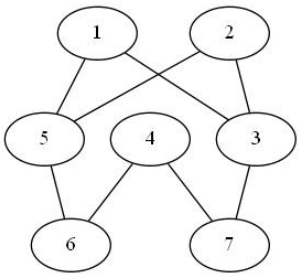


Figure 1.1: Example of a simple graph

Before considering some particular types of graph, we have to introduce some terms that regard the vertices and the edges. We say that a vertex v is **incident** with an edge e if $v \in e$; in this case e is an edge **at** v. The two vertices incident with an edge are its **ends** and an edge **joins** its ends. We will usually denote an edge $\{x, y\}$ with the more compact form xy (or yx). Two vertices x, y in G are **adjacent** (or **neighbors**) if xy is an edge of G. Two edges $e \neq f$ are adjacent if they have an end in common.

We can now define some types of graph:

- A graph in which all the vertices are pairwise adjacent is called **complete graph**;
- A **path** is a non-empty graph $P = (V, E)$, with $V = \{x_0, \dots x_k\}$ and $E = \{x_0x_1, x_1x_2, \dots x_{k-1}x_k\}$ where the x_i's are all distinct;
- A **cycle** is defined as a path, with the further characteristic that $x_0 = x_k$;

Figure 1.2: Example of a cycle and a path

- A graph $G' = (V', E')$ is a **subgraph** of $G = (V, E)$ if $V' \subseteq V$ and $E' \subseteq E$;
- If any two of the vertices of a non-empty graph are linked by a path, we say that the graph is **connected**;
- A maximal connected subgraph of G is called a **component** of G.[1] We say that two vertices that belong to the same component are **connected**;

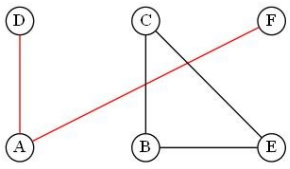


Figure 1.3: Example of a graph with two components

- A **tree** T is a connected graph which contains no cycles. The vertices with one only neighbor are the **leaves** of the tree.

Any tree has the following two properties:

[1]The term maximal means that it is the largest possible subgraph: you could not find another node anywhere in the graph such that it could be added to the subgraph and all the nodes in the subgraph would still be connected.

- Any two vertices of T are linked by a unique path in T;
- A connected graph with n vertices is a tree if and only if it has $n-1$ edges.

In the following, we will consider a particular class of trees, in which a vertex has a special role and we will call this vertex **root**. In this case we can refer to this graph as **rooted tree**. In this kind of trees is present a sense of directionality: the graph starts at its root and then it develops through the paths joining the root to the leaves.[2]

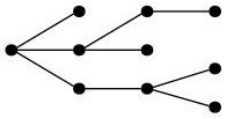


Figure 1.4: Example of a rooted tree

1.2 Random graphs

Roughly speaking, we can define a random graph to be a graph which is generated by some random process. The notion of random graph finds its origin in the fifties and is deeply linked to the paper of Paul Erdős and Alfréd Rényi "*On the evolution of random graphs*". The model introduced by the Hungarian mathematicians is a very easy one and can be described as follows: consider a graph G, whose vertex set, $V(G)$, is such that $|V(G)| = n$. Then a graph is chosen at random from the set of all $2^{\binom{n}{2}}$ possible graphs with vertex set size equal to n, where every possible graph has an equal probability of being chosen. We can be a little more formal, saying that this model can be described as the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the set of all graphs with vertex set size equal to n, $\mathcal{F}$ is the family of all subsets of Ω and for every graph $\omega \in \Omega$ the probability of being chosen, $\mathbb{P}(\omega)$, is equal to $2^{-\binom{n}{2}}$. The two most common models of random graphs used nowadays, the *binomial model* and the *uniform model*, both originate from this simple model. All the contents of this section are based on [3].

The binomial model

In this model we consider a graph with n vertices in which an edge joining any pair of vertices is drawn, independently from the others, with a certain probability p. It can be viewed as the result of $\binom{n}{2}$ independent coin tosses with a probability of success p: that is, it depends on the result of a binomial distribution with parameters $\left(\binom{n}{2}, p\right)$.

Definition 1.1. *Given a real number $p \in [0,1]$, the **binomial random graph**, denoted by $G(n,p)$ is defined by taking as Ω the set of all graphs with n vertices*

[2] In the context of rooted trees, the definition of leaves that has given may be misleading. In fact the root vertex may be connected to only one other vertex: however, we will not consider in this situation the root as a leaf.

and setting

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

The main advantage of the binomial model is the independence of presence of edges, but the disadvantage is that the number of edges is not fixed.

The uniform model

This other model differs from the previous one basically in the fact that the size of the edge set is fixed to be equal to a constant M. The probability distribution of a graph with n vertices and M edges to be chosen is uniform.

Definition 1.2. *Given an integer $M \in [0, \binom{n}{2}]$, the **uniform random graph**, denoted by $G(n, M)$, is defined by taking as Ω the family of all graphs with n vertices with exactly M edges, and as $\mathbb{P}$ the uniform probability on Ω, that is*

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

We can notice that the two models introduced are related, in the sense that, if one conditions the binomial model on the event $|E(\mathbb{G}(n,p))| = M$, then a uniform space, that characterized the uniform model, is obtained. Moreover, it is possible to prove that in many cases the two models are asymptotically equivalent, provided that $\binom{n}{2}p$, which corresponds to the expected number of edges in the binomial model, is close to M.

Random graph processes

In order to give an initial idea, we can say that a random graph process is a stochastic process that describes a random graph evolving in time. It is a family $\{G(t)\}_{t\in T}$ [3] of random graphs, with $T \subseteq \mathbb{R}^+$, where the parameter t is interpreted as time that can be either discrete or continuous. We will consider processes of random graphs with a fixed vertex set, starting without any edges and growing monotonically by adding edges according to a certain rule, but never deleting any of them. In this context, an important process is the one with the following dynamics (sometimes called *the random graph process*): the process begins with no edges at time 0 and adds new edges, one at a time, according to a uniform probability among the edges that have not been selected yet. Hence this random graph process is a Markov process, since the distribution of the future edge selection depends exclusively on the actual configuration of the graph. The time takes integer value in the set $\left(0, 1, \ldots, \binom{n}{2}\right)$, and the M-th stage of the process can be identified with the uniform random graph $G(n, M)$.

1.3 Probabilistic methods

In this third section we introduce the concepts of probability theory that are necessary to develop the analysis of random graphs. We now introduce some notations too. If a random variable X is a binomial random variable with

[3] With this nomenclature, we are not specifying any particular model for the random graphs considered.

parameters (n, p), we will denote it as $X \sim B(n, p)$, while if Y is a Poisson with mean λ, we will write $Y \sim P(\lambda)$. Moreover, if X is a Bernoulli with mean p, we will simply write $X \sim B(p)$ and if Y is a continuous uniform in the interval $[a, b]$, then $Y \sim U[a, b]$. Finally, we say that a sequence of independent and identically distributed random variables are i.i.d. random variables and a sequence of i.i.d. random variables following a certain distribution will be denoted with $\overset{i.i.d.}{\sim}$.

1.3.1 Inequalities and large deviation theory

At the beginning of this section, we recall two of the most famous inequalities in probability theory: Markov and Chebyshev inequalities. They will become very handy during the analysis of the phase transition and, for this reason, they are now introduced. For all the inequalities, let X be a non-negative random variable with $E(X) < \infty$, then, for any $t > 0$, we have that

$$\textbf{Markov: } P(X \geq t) \leq \frac{E(X)}{t}, \tag{1.1}$$

$$\textbf{Chebyshev: } E(|X - E(X)| \geq t) \leq \frac{Var(X)}{t^2}. \tag{1.2}$$

Note that from (1.2) can be derived that

$$P(X = 0) \leq \frac{Var(X)}{E(X)^2}. \tag{1.3}$$

At the end of this section we will introduce another inequality, which is related to large deviation theory: an important part of modern probability that we are going to introduce briefly.

Large deviation theory

The theory of large deviations is a part of probability theory that concerns the asymptotic behavior of tails of sequences of probability distributions: that is, when a sum of random variables deviates from its mean by more than a "normal" amount. In order to understand what is meant by "normal" amount, let's consider a sequence of i.i.d. random variables $X_1, X_2, \ldots X_n$, with means μ and variances σ^2, and consider $S_n = X_1 + \ldots X_n$. The famous central limit theorem (CLT) states that

$$\frac{1}{\sqrt{n\sigma^2}}(S_n - n\mu) \underset{n\to\infty}{\to} Normal(0, 1).$$

The interpretation is that the CLT quantifies the probability that S_n differs from its expectation by an amount of order $\sqrt{n}$: deviations of this size are called *normal*. A very gentle introduction about this area can be found in the bibliographic reference [4]. The central point about large deviation theory is to quantify the rate at which S_n differs from its expectation by an amount of order n (which is a *large* amount, because it is beyond what stated by the CLT), as $n \to \infty$. It can be proven that this decay happens at an exponential rate in n. This result is expressed in the following theorem, which will not be proven, since it is not instructive for the sequel.

Theorem 1.3. *Let $X_1, X_2, \ldots$ be i.i.d. random variables with values in $\mathbb{R}$, satisfying*

$$\phi(t) = E(e^{tX_1}) < \infty, \quad \forall t \in \mathbb{R},$$

and let $S_n = \sum_{i=1}^{n} X_i$. Then, for all $a > E(X_1)$,

$$\lim_{n\to\infty} \frac{1}{n} \ln\left(P(S_n \geq an)\right) = -I(a), \tag{1.4}$$

where

$$I(a) = \sup_{t\in\mathbb{R}} [at - \ln(\phi(t))]. \tag{1.5}$$

From this theorem becomes apparent how the decay of the tails of probability distributions is exponential in n, because of the presence of the logarithmic transformation in (1.4). The main interesting part of the theorem for the analysis of random graphs is the so-called *rate function* $I(a)$, that summarizes the behavior of tails of distributions. We now compute the rate function for some distributions of interest.

Example 1.4. ***Consider $X \sim B(p)$, with $p \in [0,1]$. We first obtain $\phi_X(t)$.***

$$\begin{aligned} \phi_X(t) = E(e^{tX}) &= e^0 P(X=0) + e^t P(X=1) \\ &= 1 - p + pe^t. \end{aligned}$$

So now we have to find the supremum in t of

$$L(a) = at - \ln(1 - p + pe^t).$$

In order to do that, we have to compute the first derivative of $L(a)$ and to find the value of t for which it is equal to zero.

$$\begin{aligned} \frac{d}{dt}L(a) = a - \frac{1}{1-p+pe^t}pe^t &= 0 \\ a - ap + ape^t - pe^t &= 0 \\ pe^t(a-1) &= a(p-1) \\ e^t &= \frac{a}{p}\frac{1-p}{1-a}, \\ \Rightarrow t = \ln\left(\frac{a}{p}\right) - \ln\left(\frac{1-a}{1-p}\right). & \end{aligned}$$

We should check the second derivative in order to assure that this value of t is a supremum: however, it can be easily done and so it is omitted. We have now to substitute the maximized value of t in order to obtain $I(a)$.

$$\begin{aligned} I(a) &= a\ln\left(\frac{a}{p}\right) - a\ln\left(\frac{1-a}{1-p}\right) - \ln\left(1 - p + pe^{\ln\left(\frac{a}{p}\right) - \ln\left(\frac{1-a}{1-p}\right)}\right) \\ &= a\ln\left(\frac{a}{p}\right) - a\ln\left(\frac{1-a}{1-p}\right) - \ln\left(1 - p + p\frac{a(1-p)}{p(1-a)}\right) \\ &= a\ln\left(\frac{a}{p}\right) - a\ln\left(\frac{1-a}{1-p}\right) - \ln\left(\frac{1-a-p+ap+a-ap}{1-a}\right) \\ &= a\ln\left(\frac{a}{p}\right) - a\ln\left(\frac{1-a}{1-p}\right) + \ln\left(\frac{1-a}{1-p}\right) \\ &= a\ln\left(\frac{a}{p}\right) + (1-a)\ln\left(\frac{1-a}{1-p}\right). \end{aligned} \tag{1.6}$$

Example 1.5. ***In this second example we will do the same calculations as in the previous one, but in this case we are considering $Y \sim P(\lambda)$.***

$$\begin{aligned} \phi_Y(t) &= \sum_{y=0}^{\infty} \frac{e^{ty}e^{-\lambda}\lambda^y}{y!} \\ &= e^{-\lambda}\sum_{y=0}^{\infty} \frac{\left(\lambda e^t\right)^y}{y!} \\ &= e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}. \end{aligned}$$

In this case we have to find the supremum of

$$at - \ln\left(e^{\lambda(e^t-1)}\right),$$

which, as before, can be easily done by computing the first derivative. Indeed,

$$\begin{aligned}\frac{d}{dt}(at - \lambda(e^t-1)) = a - \lambda e^t &= 0\\ e^t &= \frac{a}{\lambda},\\ \Rightarrow t &= \ln\left(\frac{a}{\lambda}\right).\end{aligned}$$

We now substitute as before to obtain that

$$\begin{aligned}I(a) &= a\ln\left(\frac{a}{\lambda}\right) - \lambda(e^{\ln\left(\frac{a}{\lambda}\right)} - 1)\\ &= a\ln\left(\frac{a}{\lambda}\right) - \lambda\left(\frac{a}{\lambda}\right)\\ &= \lambda - a - a\ln\left(\frac{\lambda}{a}\right).\end{aligned} \tag{1.7}$$

We now relate the rate functions of the Bernoulli and the Poisson distributions. The following is based on [5].

Proposition 1.6. *Let $I_B(a)$ be the rate function of a Bernoulli random variable with parameter p as in (1.6), and let $I_P(a)$ be the the rate function of a Poisson mean p as in (1.7). Then*

$$I_P(a) \leq I_B(a).$$

Moreover, $I_P(a) > 0$, if $a \neq p$.

Proof. We first prove that the rate function of the binomial is bigger than the one of the Poisson. Recall that, by Taylor expansion, we have that

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!},$$

from which it follows that

$$e^x \geq 1 + x. \tag{1.8}$$

So, we can also consider $x = p(e^t - 1)$, so that

$$e^{p(e^t-1)} \geq 1 + p(e^t - 1) = 1 - p + pe^t,$$

which is the moment generating function of a Bernoulli. Finally, note that

$$\begin{aligned}I_B(a) &= \sup_{t\in\mathbb{R}}\left(at - \ln\left(1 - p + pe^t\right)\right)\\ &\geq \sup_{t\in\mathbb{R}}\left(at - p\left(e^t - 1\right)\right)\\ &= \sup_{t\in\mathbb{R}}\left(at - \ln\left(e^{p\left(e^t-1\right)}\right)\right)\\ &= I_P(a).\end{aligned}$$

To prove that the rate function of the Poisson distribution is bigger than zero, note that $I_P(p) = 0$ and that $\frac{d}{da}I_P(a) = \ln(a) - \ln(p)$. Thus, if $a < p$, the function $I_P(a)$ decreases, while, if $a > p$, $I_P(a)$ increases: the result easily follows from this observation. □

We are now ready to introduce the last inequality that will be fundamental during the analysis of the component size of random graphs. In the literature, it can be found with numerous different nomenclatures: however, we will refer to this inequality with the name of Chernoff bound.

Theorem 1.7. *Let $X_1, \dots, X_n$ be i.i.d. random variables with finite mean. Then, for all $a \geq E(X_1)$,*

$$P\left(\sum_{i=1}^{n} X_i \geq na\right) \leq e^{-nI(a)} \tag{1.9}$$

where $I(a)$ is the rate function of X_1. Moreover, for any $a \leq E(X_1)$

$$P\left(\sum_{i=1}^{n} X_i \leq na\right) \leq e^{-nI(a)} \tag{1.10}$$

Proof. We start proving equation (1.9). Since the exponential is an invertible function and by Markov inequality, we have that, for $t \in \mathbb{R}$

$$\begin{aligned}
P\left(\sum_{i=0}^{n} X_i \geq na\right) &= P\left(e^{t\sum_{i=0}^{n} X_i} \geq e^{tna}\right) \\
&\leq e^{-nta} E\left(e^{t\sum_{i=0}^{n} X_i}\right) \\
&= e^{-nta} E\left(e^{tX_1+\cdots+tX_n}\right) \\
&= e^{-nta} E\left(e^{tX_1}\right) \cdots E\left(e^{tX_n}\right) \\
&= e^{-nta} E\left(e^{tX_1}\right)^n \\
&= \left(e^{-ta} E\left(e^{tX_1}\right)\right)^n \\
&= e^{-n\left(ta - \ln\left(E\left(e^{tX_1}\right)\right)\right)}
\end{aligned}$$

This inequality is true for any $t \in \mathbb{R}$, and in particular it is also true for the supremum in t of the exponent of the last expression. Indeed

$$P\left(\sum_{i=0}^{n} X_i \geq na\right) \leq e^{-n\sup_{t\in\mathbb{R}}\left(ta - \ln\left(E\left(e^{tX_1}\right)\right)\right)} = e^{-nI(a)}$$

For the proof of (1.10) we can just replace X_i by $-X_i$ and repeat the same procedure. □

Note that this new inequality is simply an "optimized" version of Markov inequality for sums of random variables. Indeed, after applying the exponential transformation to the event that the sum of a sequence of random variables exceeds its mean, we have just used Markov inequality and optimized it by taking the supremum. Note also that this result, together with Proposition 1.6, implies that the right tale of a binomial distribution is thinner than the one of a Poisson, with the corresponding parameter.

1.3.2 Couplings and Poisson approximation

Couplings are a powerful probabilistic tool and find their application in numerous fields. For example, they play a fundamental role in bounding the total variation distance between two random variables, aspect that is extensively discussed in [6]. On the contrary, we will use them in this dissertation to prove the so-called Poisson approximation. A very comprehensive introduction to the topic, with numerous applications, can also be found in [7]. We have now to define what a coupling between random variables is.

Definition 1.8. *A coupling of a collection of random variables X_i, $i \in \mathbb{I}$, where $\mathbb{I}$ is some index set, is a family of random variables $(\hat{X}_i : i \in \mathbb{I})$ defined on a single probability space, such that X_i and $\hat{X}_i$ have the same distribution, $\forall i \in \mathbb{I}$.*

For example, consider $(\hat{X}, \hat{Y})$ a coupling of the random variables X and Y: the key point of the definition above is that while X and Y may be defined on different probability spaces, the coupled random variables $(\hat{X}, \hat{Y})$ are defined on the same probability space. Clearly, the easiest coupling is obtained when X and Y are independent. Thus, a coupling has a fixed marginal distribution (which is the distribution of the X_i's), and the trick is to find a joint distribution that fits one's purposes.

Example 1.9. ***Let's consider two Bernoulli random variables X and Y, with parameter $1/2$. Recall that, for example, the Bernoulli distribution describe the probability of the toss of a fair coin. One way to couple this two distributions is to define $(\hat{X}, \hat{Y})$ to be a pair of independent tosses, so that $P(X = x, Y = y) = 1/4$ for all $x, y \in \{0, 1\}$. Another way to couple X and Y is to let X being a fair coin toss and define $Y = X$: Y is forced to be equal to X. In this case, $P(X = Y = 0) = 1/2$, $P(X = Y = 1) = 1/2$ and $P(X \neq Y) = 0$.***

We now introduce a typology of couplings that will be used in the following, called *maximal coupling*: it is a coupling such that the variables coincide maximally, that is with the highest possible probability.

Theorem 1.10 (**Maximal Coupling**). *Let X_i, $i \in \mathbb{I}$, be a collection of discrete random variables taking values in a countable set Ξ, each of them with a probability mass function p_i. Then, there exists a **maximal** coupling $(\hat{X}_i : i \in \mathbb{I})$, that is a coupling such that*

$$P(C) = \sum_{x \in \Xi} \inf_{i \in \mathbb{I}} p_i(x), \tag{1.11}$$

where C is an event such that, if C occurs, then all the $\hat{X}_i$'s coincide, that is

$$C \subseteq \{\hat{X}_i = \hat{X}_j : \forall i, j \in \mathbb{I}\}.$$

Proof. The first step of the proof is to show that the probability of the maximal coupling event is less or equal to $\sum_{x \in \Xi} \inf_{i \in \mathbb{I}} p_i(x)$. In order to do that, note that for all $i, j \in \mathbb{I}$ and $x \in \Xi$, we have that

$$P(\hat{X}_i = x, C) = P(\hat{X}_j = x, C) \leq p_j(x),$$

since coupled variables have the same probability mass function of the original ones and since the probability of an intersection of events is always less or equal to the probability of one of the single events. Thus

$$P(\hat{X}_i = x, C) \leq \inf_{j \in \mathbb{I}} p_j(x), \qquad \forall i \in \mathbb{I}, x \in \Xi.$$

If we sum over $x \in \Xi$, we obtain that

$$P(C) \leq \sum_{x \in \Xi} \inf_{i \in \mathbb{I}} p_i(x).$$

We now have to construct a coupling such that the above inequality holds as an identity. Consider $c = \sum_{x \in \Xi} \inf_{i \in \mathbb{I}} p_i(x) \in (0,1)$ and let I, V, W_i be independent random variables for $i \in \mathbb{I}$ such that:

$$\begin{aligned} I &\sim B(c), \\ P(V = x) &= \frac{\inf_{i \in \mathbb{I}} p_i(x)}{c}, \qquad \forall x \in \Xi, \\ P(W_i = x) &= \frac{p_i(x) - cP(V = x)}{1 - c}, \qquad \forall x \in \Xi, i \in \mathbb{I}. \end{aligned}$$

Then, we define for each $i \in \mathbb{I}$

$$\hat{X}_i = \begin{cases} V & \text{if } I = 1 \\ W_i & \text{if } I = 0. \end{cases}$$

So, $(\hat{X}_i : i \in \mathbb{I})$ is a coupling, since

$$\begin{aligned} P(\hat{X}_i = x) &= P(I = 0)P(W_i = x) + P(I = 1)P(V = x) \\ &= (1 - c)\frac{p_i(x) - cP(V = x)}{1 - c} + cP(V = x) \\ &= p_i(x) = P(X_i = x) \end{aligned}$$

If we take $C = \{I = 1\}$, then $X_i = V \ \forall i \in \mathbb{I}$ and so it is an appropriate event. Moreover, $P(C) = c$, as required.

We now have to consider the two cases in which $c = 0$ or $c = 1$. If $c = 0$, then

$$\sum_{x \in \Xi} \inf_{i \in \mathbb{I}} p_i(x) = 0 \quad \Longrightarrow \quad \inf_{i \in \mathbb{I}} p_i(x) = 0 \text{ for every } x \in \Xi$$

and so there has to exist at least one $i \in \mathbb{I}$ for which $P(\hat{X}_i = x) = 0$ for every $x \in \Xi$. For this reason, the only possible choice is to take $C = \emptyset$ and the $\hat{X}_i$'s all independent of each other (so to obtain a coupling). On the contrary, if $c = 1$, then

$$\sum_{x \in \Xi} \inf_{i \in \mathbb{I}} p_i(x) = 1 \quad \Longrightarrow \quad \inf_{i \in \mathbb{I}} p_i(x) \text{ is a probability distribution,}$$

but this happens only if all the $\hat{X}_i$'s are identically distributed and C is the set of all possible outcomes. □

Poisson approximation

During the exploration of a component of a random graph, we will use many times the fact that, when n is large, a binomial distribution can be approximated by a Poisson distribution. We now give a coupling representation of this property.

Theorem 1.11. *Let $X \sim B(n,p)$ and let $Y \sim P(\lambda)$, where $\lambda = np$ and $p \in [0,1]$. Then, there exists a coupling $(\hat{X}, \hat{Y})$ of X and Y such that*

$$P(\hat{X} \neq \hat{Y}) \leq \frac{\lambda^2}{n}.$$

Before starting the proof, a little remark: note that Theorem 1.11 states that the probability that a binomial is different from a Poisson in the limit (that is, as the number of Bernoulli random variables that compounds the binomial gets larger) tends to zero.

Proof. Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} B(p)$, and thus $X = X_1 + \cdots + X_n$. Consider also $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} P(p)$: thus also $Y = Y_1 + \cdots Y_n$. Let $(\hat{X}_1, \hat{Y}_1), \ldots, (\hat{X}_n, \hat{Y}_n)$ be independent pairs such that, for each i, $(\hat{X}_i, \hat{Y}_i)$ is a maximal coupling of X_i and Y_i, where the existence of a maximal coupling is guaranteed by Theorem 1.10. Now set

$$\hat{X} = \hat{X}_1 + \cdots \hat{X}_n \qquad \text{and} \qquad \hat{Y} = \hat{Y}_1 + \cdots + \hat{Y}_n.$$

Clearly, $(\hat{X}, \hat{Y})$ is a coupling of X and Y. Note that

$$P(\hat{X} \neq \hat{Y}) \leq \sum_{i=1}^{n} P(\hat{X}_i \neq \hat{Y}_i).$$

We have now to deduce how to construct the maximal coupling $(\hat{X}_i, \hat{Y}_i)$: note that in this case, following the nomenclature of Theorem 1.10, we can choose the set C to be equal to $\{\hat{X}_i = \hat{Y}_i\}$ and that (1.11) can be written as

$$\begin{aligned} P(\hat{X}_i = \hat{Y}_i) &= \sum_{x=0}^{\infty} \min\{P(\hat{X}_i = x), P(\hat{Y}_i = x)\} \\ &= \min\{P(\hat{X}_i = 0), P(\hat{Y}_i = 0)\} + \min\{P(\hat{X}_i = 1), P(\hat{Y}_i = 1)\}, \end{aligned}$$

since for a Bernoulli random variable $P(X_i = x) = 0$, $\forall x \geq 2$. Recalling the result in (1.8), we can deduce that

$$P(\hat{X}_i = 0) = 1 - p \leq e^{-p} = P(\hat{Y}_i = 0),$$
$$P(\hat{X}_i = 1) = p \geq pe^{-p} = P(\hat{Y}_i = 1),$$

where the second inequality holds, since $e^{-p} \in [1/e, 1]$ for $p \in [0,1]$. Thus

$$P(\hat{X}_i = \hat{Y}_i) = 1 - p + pe^{-p} \geq 1 - p + p(1-p) = 1 - p^2$$
$$\implies P(\hat{X}_i \neq \hat{Y}_i) \leq p^2.$$

Finally,

$$\begin{aligned} P(\hat{X} \neq \hat{Y}) &\leq \sum_{i=1}^{n} P(\hat{X}_i \neq \hat{Y}_i) \\ &\leq \sum_{i=1}^{n} p^2 = \sum_{i=1}^{n} \left(\frac{\lambda}{n}\right)^2 \\ &= n\frac{\lambda^2}{n^2} = \frac{\lambda^2}{n}. \end{aligned}$$

□

1.3.3 Stochastic domination

The concept of stochastic order is widely used in probability and decision theory in order to compare random variables. It will play a fundamental role in the third chapter and, for this reason, a brief introduction is necessary. This section is mainly based on [7]. We start by defining what is meant by the term *stochastic domination*.

Definition 1.12. *Let X and Y be two $\mathbb{R}$-valued random variables. We say that X is stochastically dominated by Y if, for every $x \in \mathbb{R}$,*

$$F_X(x) \geq F_Y(x),$$

where F is the the distribution function of the random variables. We denote this by $X \preceq Y$.

This notion, roughly speaking, relates two random variables by considering the chance of one being "bigger" than one other. It is somehow a probabilistic equivalent of the pointwise domination, which is the classical $>$ or $<$ relationship between functions. Consider two functions f and g, defined on the whole real line: we say that f is pointwise dominated by g if $f(x) < g(x)$, $\forall x \in \mathbb{R}$. However, we can relate the concept of pointwise domination to probability theory by simply noting that, X is pointwise dominated by Y, if $P(X \leq Y) = 1$, where X and Y are $\mathbb{R}$-valued random variables.

Now we introduce a coupling characterization of stochastic domination, which will be used later.

Theorem 1.13. *Let X and Y be two real-valued random variables. Then X is stochastically dominated by Y if and only if there exists a coupling $(\hat{X}, \hat{Y})$ of X and Y such that $\hat{X} \leq \hat{Y}$: that is, $\hat{X}$ is pointwise dominated by $\hat{Y}$.*

Proof. If $\hat{X} \leq \hat{Y}$, then $\{\hat{Y} \leq x\} \subseteq \{\hat{X} \leq x\}$, which implies that $P(\hat{Y} \leq x) \leq P(\hat{X} \leq x)$ and so

$$F_Y(x) \leq F_X(x), \qquad \forall x \in \mathbb{R}.$$

For the other direction, note that if X is stochastically dominated by Y, $F_X(x) \geq F_Y(x)$, $\forall x \in \mathbb{R}$. It follows that, for all $u \in [0, 1]$,

$$F_X^{-1}(u) \leq F_Y^{-1}(u),$$

where F^{-1} is, in the discrete case, the generalized inverse of F, defined as

$$F_X^{-1}(u) = \{\inf x \in \mathbb{R} : F_X(x) \geq u\}.$$

However, it is well-known that, if $U \sim U[0, 1]$, then

$$X = F_X^{-1}(U) \qquad \text{and} \qquad Y = F_Y^{-1}(U),$$

from which the result follows. □

There are several natural examples of pairs of random variables that are stochastically ordered, and we will see two of them in the following examples.

Example 1.14. ***Let's consider two Poisson random variables,*** $X \sim P(\lambda)$ ***and*** $Y \sim P(\mu)$***, where*** $\lambda \leq \mu$***. We can construct a coupling of these two random variables by simply letting*** $\hat{X} \sim P(\lambda)$ ***and*** $\hat{Z} \sim P(\mu - \lambda)$***, where*** $\hat{X}$ ***and*** $\hat{Z}$ ***are independent. Let also*** $\hat{Y} = \hat{X} + \hat{Z}$ ***and thus*** $\hat{Y} \sim P(\mu)$***, since the sum of independent Poisson random variables is a Poisson random variable too. Moreover, since all the distributions are bigger or equal than zero, we have that*** $\hat{X} \leq \hat{Y}$***, which means that, because of Theorem 1.13,*** $X \preceq Y$***.***

Example 1.15. ***Consider now two binomial random variables,*** $X \sim B(m, p)$ ***and*** $Y \sim B(n, p)$***, where*** $m \leq n$ ***and*** $p \in [0, 1]$***. Also in this case we can construct a simple coupling between these two variables, by simply recalling that a binomial random variable is the sum of i.i.d. Bernoulli random variables. So, let*** $\hat{X} = \sum_{i=1}^{m} Z_i$ ***and*** $\hat{Y} = \sum_{i=1}^{n} Z_i$***, where*** $Z_1, \dots, Z_n \overset{i.i.d.}{\sim} B(p)$***. Then, since the Bernoulli is a non-negative random variable and since*** $\hat{Y} = \hat{X} + Z_{m+1} \cdots + Z_n$***, we have that*** $\hat{X} \leq \hat{Y}$ ***and thus*** $X \preceq Y$ ***.***

Note that, by the same reasoning, it can be proven that $X_1 \preceq Y$***, where in this case*** $X_1 \sim B(n - Z, p)$***, with*** Z ***any non-negative random variable.***

1.3.4 Markov chains and random walks

In the next chapter we will give a random walk construction of branching processes, and the same idea will be fundamental while exploring the components of a random graph. For this reason, a very brief introduction on the topic is necessary and we will start with the definition of a discrete Markov chain: random walks are a particular case of this stochastic process. The following is mostly based on the bibliographic reference [8].

Definition 1.16. *Let* $X_0, X_1, X_2, \dots$ *be a collection of random variables taking values in* $E \subseteq \mathbb{Z}$*. Then,* $\{X_t\}_{t=0}^{\infty}$ *is a Markov chain if,* $\forall t \geq 0$*,*

$$P(X_t = x_t | X_{t-1} = x_{t-1}, \dots, X_1 = x_1, X_0 = x_0) = P(X_t = x_t | X_{t-1} = x_{t-1}). \tag{1.12}$$

The condition in (1.12) is also known as ***Markov property*** *and the elements of* E *are called the* ***states*** *of the chain.*

If we think of the subscript t of the previous definition as an index of time, like days, the main characteristic of a Markov process is that what will happen tomorrow depends only on what is the state of the process today. For example, we can think of a frog that everyday makes a jump between lily pads. If the position in which we will find the frog tomorrow is influenced only by the lily pad in which it lies today, then we can say that the jumping frog moves according to a Markov chain.

There are many interesting properties of this kind of processes and a lot of characteristics to be studied: however, we will only introduce the so-called **hitting time**.

Definition 1.17. *Let* $\{X_t\}_{t=0}^{\infty}$ *be a Markov chain with state space* E*. Then, for* $x \in E$*, we say that* τ_x *is an hitting time if*

$$\tau_x = \min\{t \geq 0 : X_t = x\} \tag{1.13}$$

Thus, an hitting time is the first time in which the chain visits a particular state. In the following, when necessary, we will stress the position from which the chain starts, by inserting the subscript $P_k(\dots)$, if the chain starts from the state k.

We are now ready to give the definition of a random walk.

Definition 1.18. *Let $X_0, X_1, X_2, \dots$ be i.i.d. discrete random variables taking values in a set $E \subseteq \mathbb{Z}$. Let also $S_n = k + X_1 + X_2 + \cdots X_n$, where $n > 0$ and $X_0 = k$. Then, $\{S_n\}_{n=1}^{\infty}$ is a random walk starting at k.*

Clearly, a random walk is a Markov chain since $S_n = S_{n-1} + X_n$. There are a lot of examples of random walks: the simplest one is a particle that moves on the integer line. In a more general case, we can think of a particle that moves on the vertices of a graph, in which every vertex is denoted by an integer number.

The definition a random walk is almost all we need to proceed to the study of branching processes: we have to prove only one property of random walks, that will be exploited in the next two chapters. The proof of the following theorem is based on [9].

Theorem 1.19. *Let $\{S_n\}_{n=0}^{\infty}$ be a random walk starting at $k \geq 0$, such that $S_n = k + X_1 + \cdots X_n$, where $X_1, X_2, \dots$ are i.i.d. random variables satisfying $P(X_1 \geq -1) = 1$. Let also τ_0 be the hitting time of the origin. Then*

$$P_k(\tau_0 = n) = \frac{k}{n} P_k(S_n = 0). \tag{1.14}$$

Proof. We will prove this result by induction on n. We start by showing that (1.14) holds when $n = 1$. If $k = 0$, the right-hand side is equal to zero, since there is a k on the numerator, while for the left-hand side, note that $P_0(\tau_0 = 1) = 0$. If $k > 1$, both sides are equal to zero too, since the walk can make only one step when $n = 1$ and $X_1 \geq -1$ almost surely.[4] If $k = 1$, $P_1(\tau_0 = 1) = P_1(S_1 = 0) = P_1(1 + X_1 = 0) = P_1(X_1 = -1)$ and $\frac{k}{n} P_k(S_n = 0) = P_1(S_1 = 0)$: thus (1.14) holds for $n = 1$.

We suppose it is also true for $n - 1$. We can restrict the analysis to the case $k \geq 1$, since, again, for $k = 0$ both sides are equal to zero. Then, by the law of total probability

$$P_k(\tau_0 = n) = \sum_{s=-1}^{\infty} P_k(\tau_0 = n | X_1 = s) P(X_1 = s).$$

Because of the Markov property, we can also note that

$$P_k(\tau_0 = n | X_1 = s) = P_{k+s}(\tau_0 = n - 1) = \frac{k + s}{n - 1} P_{k+s}(S_{n-1} = 0),$$

where the last equality holds because of the inductive hypothesis. Thus, using

[4] Note that this observation can be easily extended for any value of n. Indeed, if $k > n$, $P_k(\tau_0 = n) = 0$, since the walk can move back by at most n in n steps.

Bayes theorem and again the Markov property:

$$\begin{aligned}
P_k(\tau_0 = n) &= \sum_{s=-1}^{\infty} \frac{k+s}{n-1} P_{k+s}(S_{n-1} = 0) P(X_1 = s) \\
&= \sum_{s=-1}^{\infty} \frac{k+s}{n-1} P_k(S_n = 0 | X_1 = s) P(X_1 = s) \\
&= \sum_{s=-1}^{\infty} \frac{k+s}{n-1} P_k(X_1 = s | S_n = 0) P_k(S_n = 0) \\
&= \frac{k}{n-1} P_k(S_n = 0) \sum_{s=-1}^{\infty} P_k(X_1 = s | S_n = 0) \\
&\quad + \frac{1}{n-1} P_k(S_n = 0) \sum_{s=-1}^{\infty} s P_k(X_1 = s | S_n = 0) \\
&= \frac{k}{n-1} P_k(S_n = 0) + \frac{1}{n-1} P_k(S_n = 0) E_k(X_1 | S_n = 0) \\
&= \frac{P_k(S_n = 0)}{n-1} (k + E_k(X_1 | S_n = 0)).
\end{aligned}$$

Now, since the X_i's are identically distributed, $E_k(X_i|S_n = 0)$ is equal for all $i = 1, \ldots, n$. Thus

$$E_k(X_1|S_n = 0) = \frac{1}{n} \sum_{i=1}^{n} E_k(X_i|S_n = 0) = \frac{1}{n} E_k \left(\sum_{i=1}^{n} X_i | S_n = 0 \right) = -\frac{k}{n},$$

since the random walk starts at k and it is forced to be equal to zero at $S_n = \sum_{i=1}^{n} X_i$. Finally,

$$\begin{aligned}
P_k(\tau_0 = n) &= \frac{P_k(S_n = 0)}{n-1} \left(k - \frac{k}{n} \right) \\
&= \frac{P_k(S_n = 0)}{n-1} \left(\frac{k(n-1)}{n} \right) \\
&= \frac{k}{n} P_k(S_n = 0)
\end{aligned}$$

as required. □

Chapter 2

Branching processes

Branching processes are a very important tool in the study of the phase transition of the largest component size of a random graph and, for this reason, we will introduce the main features of this kind of processes. In the first section we will define what a branching process is; in the second one we will study some useful characteristics, while in the third we will introduce a different characterization of this stochastic process. Finally, in the last section, we will study two particular branching processes. The whole chapter is mainly based on the bibliographic references [10] and [5].

2.1 Introduction

A branching process is a stochastic process that may be used as a simple model to describe a bacterial growth or the spread of a family name. This process models a population in which each individual in generation[1] n produces some random number of individuals in generation $n+1$, according to a fixed probability distribution that does not vary from individual to individual. We will always consider processes that start with one individual in the first generation (generation zero). The usual way to represent such a process is through a rooted tree, in which every level represents a generation.

In Figure 2.1 we can see a tree that describes the dynamic of a branching process. The first generation consists of an individual which generates three *offspring*: the first one (on the left) generates a single offspring, the second one (in the middle) has two offspring while the last one does not generate any offspring. The process can continue in this way until a generation of individuals does not produce any offspring: in this case the process dies out and becomes *extinct*. We are now ready to define this process.

Definition 2.1. *Let X be an integer-valued non-negative random variable with probability mass function $p_X(x)$, for $x = 0, 1, \dots$. We say that a sequence of random variables $\{Z_n\}_{n\in\mathbb{N}}$ is a* ***branching process*** *if*

- $Z_0 = 1$;
- $Z_{n+1} = X_1^{(n)} + X_2^{(n)} + \cdots + X_{Z_n}^{(n)}$;

[1] With the term generation we refer to the set of individuals born at the same time.

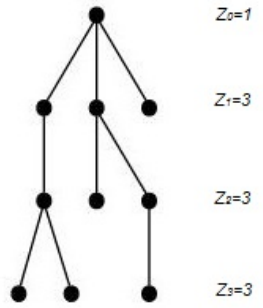


Figure 2.1: Example of a tree representing a Branching process

where all the random variables $X_j^{(n)}$'s have the same distribution as X and are independent of each other.

The generic variable $X_j^{(n)}$ represents the number of offspring of the individual j that generates this offspring at time n and it is usually called family-size distribution or offspring distribution. In Definition 2.1 we have stressed the fact that the variables X_j's correspond to the n-th generation by putting the letter n as superscript. However, in the following, the n will be dropped in the situations in which this causes no doubts, in order to make the nomenclature easier. Clearly, a branching process is a Markov chain since the number of individuals in the n-th generation depends only on the number of individuals in the previous one.

From now on we can assume that $P(X = x) \neq 1$ for $x = 0, 1, \ldots$ in order to avoid trivialities: indeed, if the probability of having k offspring is equal to one, then there is no randomness in the process.

2.2 Characteristics of the process

In this section we are going to introduce three main properties of branching processes: the *mean behavior of the process*, the *total progeny* and the *extinction probability* .

The mean behavior of the process

We want to derive the first moment of the process at time n. The first step in the study of branching processes is to understand how to find the distribution of the Z's in term of the family-size mass function. Clearly, $Z_0 = 1$ and $P(Z_1 = x) = p_X(x)$ for $x = 0, 1, \ldots$. It is not so easy, however, to derive the mass function of Z_2 since it is the sum of a random number of i.i.d. random variables. This kind of sums is treated in an easier way by using the probability generating functions. We write

$$G_n(s) = E\left(s^{Z_n}\right) = \sum_{x=0}^{\infty} s^x P(Z_n = x)$$

for the probability generating function of Z_n, and

$$G_X(s) = \sum_{x=0}^{\infty} s^x p_X(x)$$

for the probability generating function of a typical family size. Our first goal is to express G_n in terms of G_X.

Theorem 2.2. *The probability generating functions G_X, G_0, G_1, ... satisfy*

$$G_0(s) = s, \qquad G_n(s) = G_{n-1}(G_X(s)) \qquad \text{for } n = 1, 2, \ldots$$

and hence G_n is the n-th iterate of G_X:

$$G_n(s) = G_X(G_X(\cdots G_X(s)\cdots)) \qquad \text{for } n = 0, 1, \ldots$$

Proof. For the first two results, we can note that

$$G_0(s) = E\left(s^{Z_0}\right) = E(s^1) = s.$$

$$\begin{aligned} G_n(s) &= E\left(s^{Z_n}\right) = E\left(s^{X_1+\ldots X_{Z_{n-1}}}\right) \\ &= \sum_{i=0}^{\infty} E\left(s^{X_1+\ldots X_{Z_{n-1}}} | Z_{n-1} = i\right) P(Z_{n-1} = i) \\ &= \sum_{n=0}^{\infty} E\left(s^{X_1+\ldots X_i}\right) P(Z_{n-1} = i) \\ &= \sum_{n=0}^{\infty} [G_X(s)]^i P(Z_{n-1} = i) = G_{n-1}(G_X(s)). \end{aligned}$$

Hence, iterating this result

$$\begin{aligned} G_n(s) &= G_{n-1}(G_X(s)) = G_{n-2}(G_X(G_X(s))) \\ &= \cdots = G_X(G_X(\cdots G_X(s)\cdots)). \end{aligned}$$

□

We are now ready to derive the mean value of the number of individuals in a generic generation n.

Theorem 2.3. *Let μ be the mean of the family-size distribution, that is $\mu = \sum_{x=0}^{\infty} x p_X(x)$. Then*

$$E(Z_n) = \mu^n. \tag{2.1}$$

Proof. By the properties of the probability generating function, we have that

$$\begin{aligned} E(Z_n) &= G_n'(1) = G_{n-1}'(G_X(1)) G_X'(1) \\ &= G_{n-1}'(1) G_X'(1) = \mu E(Z_{n-1}) \\ &= \mu^2 E(Z_{n-2}) = \cdots = \mu^n E(Z_0) = \mu^n. \end{aligned}$$

□

Formula 2.1 gives an idea about the mean behavior of the process at the limit. Indeed

$$\lim_{n\to\infty} E(Z_n) = \begin{cases} 0 & \text{if } \mu < 0 \\ 1 & \text{if } \mu = 1 \\ \infty & \text{if } \mu > 1 \end{cases}$$

So, if μ is less than one we expect the process to become extinct, while, if μ is bigger than one, the mean number of individuals in a generation at the limit is not zero. We will show that, if $\mu > 1$, there is a certain probability for the process not to become extinct. This remark shows the presence for branching processes of a so-called phase transition. Until $\mu < 1$ there is a certain behavior, which all of a sudden changes for $\mu > 1$. This is the reason why branching processes are so important in the study of the phase transition of the largest component of a random graph, since both processes share this behavior. If $\mu = 1$, then we are in the critical case. We are now ready to derive the results for the probability of the population to become extinct.

Extinction probability

We denote the probability of ultimate extinction by

$$\eta = P(\exists n : Z_n = 0),$$

while we define the probability that the branching process is extinct by the nth generation as

$$\eta_n = P(Z_n = 0).$$

The following theorem contains every information about the extinction of a branching process.

Theorem 2.4. *The probability η of ultimate extinction is the smallest non-negative root of the equation*

$$x = G_X(x). \tag{2.2}$$

Moreover, this probability η is equal to one if and only if the mean family-size μ satisfies $\mu \leq 1$.

Remark 2.5. *We suppose that $P(X = 0) > 0$, since otherwise $\eta = 0$ and $\mu > 1$.*

Proof. The first step in this proof consists of showing that equation (2.2) is true. Firstly note that if $Z_n = 0$, then $Z_m = 0$, for all $m \geq n$. Thus, $\eta_n \leq \eta_{n+1}$, since '$Z_n = 0$' implies '$Z_{n+1} = 0$'. Hence

$$\eta = \lim_{n\to\infty} \eta_n$$

exists, since the sequence of η_n's concerns an increasing sequence of events with limit η. We also note that $G_n(0) = P(Z_n = 0) = \eta_n$. Now, since

$$\begin{aligned} G_n(s) &= G_{n-1}(G_X(s)) = G_X(G_X(\cdots(s)\cdots)) \\ &= G_X(G_{n-1}(s)), \end{aligned}$$

setting $s = 0$, we obtain that

$$\eta_n = G_X(\eta_{n-1}),$$

with boundary condition $\eta_0 = 0$.[2] Taking limits, it follows that

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

We now have to show that η is the smallest non-negative root of equation (2.2). Suppose that ϵ is any non negative root of (2.2) too. We shall show that $\eta \leq \epsilon$. Firstly, note that G_X is a non decreasing function on $[0, 1]$, since it has non negative coefficients[3]. Hence,

$$\eta_1 = G_X(0) \leq G_X(\epsilon) = \epsilon$$

because $\epsilon \in [0, 1]$. Then

$$\eta_1 \leq \epsilon \qquad \Longrightarrow \qquad G_X(\eta_1) \leq G_X(\epsilon),$$

and so

$$\eta_2 \leq \epsilon.$$

Iterating this process, it follows that

$$\eta_n \leq \epsilon \qquad \text{for } n = 1, 2, \ldots$$

from which it follows that

$$\eta = \lim_{n\to\infty} \eta_n \leq \epsilon.$$

At this point, we just have to prove that the probability of ultimate extinction is equal to one if and only if the mean family-size is less or equal than one. We have already seen and, implicitly used, the fact that in $[0, 1]$ the function G_X is continuous and non-decreasing. It is also possible to show that it is convex (by simply computing the second-derivative of the series and noting that it is always positive). For these three reasons the shape of the curve G_X may look like the ones in Figure 2.2. By looking at the picture, we deduce that there can be either one or two intersection in $[0, 1]$ between the curve $y = G_X(x)$ (black line) and the red line corresponding to $y = x$. The smallest intersection between these two

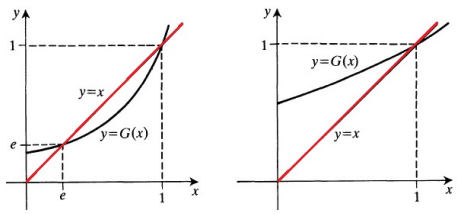


Figure 2.2: The two possible shapes of G_X

curves, corresponds to the probability of ultimate extinction. So, in the graph on the left we notice that this probability is less than one, because there are two distinct intersections, while in the other one there is a unique intersection at $x = 1$, and so $\eta = 1$. However, we can also notice in the graph on the left that the slope of the curve $G_X(x)$ at $x = 1$ is bigger than one (which corresponds to the slope of the straight line), concluding that in this case $\mu = G_X^{'}(1) > 1$. By same reasoning, we conclude that $\eta = 1$ if and only if $\mu = G_X^{'}(1) \leq 1$. □

[2]This follows from the fact that at time zero, there is certainly one and only one individual.

[3]The coefficients of the sum $G_X = \sum_{x=0}^{\infty} s^x P(X = x)$ are the components of the probability mass function, that is, they are probabilities: hence bigger than zero.

Total progeny

Definition 2.6. *Let $\{Z_n\}_{n\in\mathbb{N}}$ be a branching process. Then*

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

is called the total progeny of the branching process.

Clearly, the total progeny represents the number of all the individuals in a branching process: this number is finite if the process becomes extinct. It will be fundamental to quantify the size of a component of a random graph: indeed, we will compare a component of a graph to a branching process, and its size will be described by the total progeny.
We now consider the expected total progeny.

Proposition 2.7. *For a branching process $\{Z_n\}_{n\in\mathbb{N}}$, in which the expected family-size μ is less than one, the expected total progeny is*

$$E(T) = \frac{1}{1-\mu}. \tag{2.3}$$

Proof.

$$E(T) = E\left(\sum_{n=0}^{\infty} Z_n\right) = \sum_{n=0}^{\infty} E(Z_n) = \sum_{n=0}^{\infty} \mu^n = \frac{1}{1-\mu},$$

where the possibility of interchanging the summation and the expectation is guaranteed by monotone convergence. □

2.3 Random walk perspective

In the previous sections of this chapter we have always considered branching processes by looking at the number of descendant of each generation. We can also look at them from another perspective, which will be really useful for the analysis of random graphs: we can sequentially investigate the number of children of each member of the population. In the perspective introduced in the first section, we look at a whole generation at a time, starting from the root vertex to the successive generations. In the random walk perspective we do not care about the generations as a whole, but we just look at the offspring of every single individual. At every time step, we look at an individual and we discover its number of offspring. Then we pick any other vertex, in any generation, we look at its progeny and so on. This way of considering branching processes leads to the *random walk formulation*. The reason for this terminology will become apparent later and we now start by giving a simple representation of it. Consider a branching process and consider the root vertex as the only *active* individual. At the generic time i, we select one of the active individuals in the population, and give it X_i children, where X_i is, as before, the random variable that represents the family-size. The children (if any) of the selected individual are set to active, and the individual becomes *inactive*. This process continues as long as there are active individuals in the population. This means that if the process never dies out, we will keep looking for new active individuals an

infinite number of times. On the contrary, if the population dies out, once all the individuals that correspond to leaf vertices are inactive, there will be no more active individuals to investigate. Intuitively, the set of active individuals corresponds to the number of individuals we have seen but whose children we have not yet explored.

There are two canonical ways in which we can sequentially choose which is the vertex to look at. These two procedures correspond to two well-known algorithms: the **depth-first search** and the **breadth-first search**. In order to explain this algorithms we need first to introduce a labeling for trees, which is known as **Ulam-Harris labeling**.

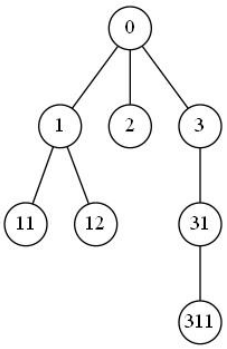


Figure 2.3: Example of a tree with Ulam-Harris labeling

The labeling works in the following way: we assign a conventional symbol to the root vertex (in Figure 2.3 is 0, but it can also be $\emptyset$); then, we assign to the first generation numbers from 1 to n (from the left to the right or vice versa), where n is the number of individuals in the first generation. Then, in the second generation every individual is identified by two numbers: the first one corresponds to the number of the adjacent vertex in the first generation, while the second one is assigned, as before, among all the progeny of the adjacent vertex considered. So, for example, in Figure 2.3, there are three vertices in the second generation: the first one on the left has label 11 because is the first children of the individual labeled 1. The process of labeling continues in this fashion as long as there are vertices to be labeled.

Now that we have introduced this nomenclature we can explain how the breadth-first search and the depth-first search work, with the help of a simple example. A very precise discussion about the algorithms can be found in [11], where they are introduced from a computational point of view.

The **depth-first search** algorithm, once a vertex is visited, explores all its progeny, going deep down the tree. In this way it is possible that vertices far from the root are visited before vertices that belong to the first generation. For example, if we apply this algorithm to the tree in Figure 2.3 the lexicographical order in which we will explore the vertices is 0, 1, 11, 12, 2, 3, 31, 311.

On the contrary, in the **breadth-first search** algorithm the vertices are explored according to their generation: after the root vertex, it visits all the vertices of the first generation, then the second one and so on. This algorithm clearly investigates the branching process in the classic way, by looking at a

whole generation before moving to others. In this other case the vertices of the tree in Figure 2.3 are explored with the following lexicographical order: 0, 1, 2, 3, 11, 12, 31, 311.

A way in which is easy to describe the "search" process is through a dot-plot, in which the x-axis represents the number of the iteration, while the y-axis represents the number of active individuals after the corresponding iteration. In Figures 2.4 and 2.5 we can see the dot-plots corresponding to the search

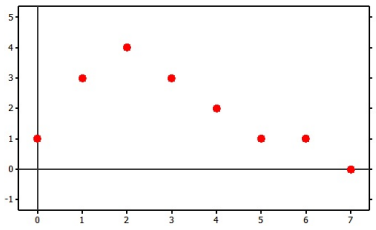


Figure 2.4: Number of active individuals in a depth-first search

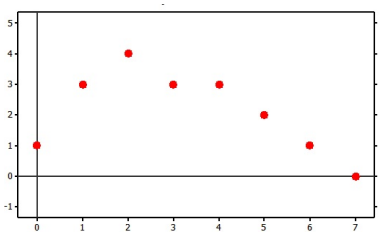


Figure 2.5: Number of active individuals in a breadth-first search

process with the two algorithms for the branching process described by the tree in Figure 2.3.

We are now ready to introduce the random walk representation of a branching process. Let $X_1, X_2, \dots$ be a sequence of i.i.d. random variables, each of them representing the family-size of an individual. We define $S_0, S_1, \dots$ by the recursion:

$$S_0 = 1 \tag{2.4}$$

$$S_i = S_{i-1} + X_i - 1 = X_1 + \cdots + X_i - (i-1). \tag{2.5}$$

Let also T be the smallest t for which $S_t = 0$, that is

$$T = \inf\{t : S_t = 0\} = \inf\{t : X_1 + \cdots X_t = t - 1\}. \tag{2.6}$$

S_i represents the number of active individuals after the i-th investigation, that is, the number of seen, but not yet explored, individuals, as we noted earlier. It is also clear that, the sequence of S_i's represents a stochastic process really close to a random walk: indeed, it is simply the sum of some i.i.d. random variables, in which the value of the process at time i is determined only by the position of the process at time $i-1$ and by the value of the variable X_i (there is also

a correction term $i-1$ that subtracts the number of inactive individuals). We can actually relate equations (2.4) and (2.5) to Definition 1.18 of a random walk by simply considering $S_i = S_{i-1} + Y_i$, where $Y_i = X_i - 1$, $\forall i \geq 1$ and $S_0 = 1$. We can also note that T represents the total progeny of the process, because we defined it to be the first time (actually it is also the only one) in which there are no more active individuals to explore. This description is equivalent to Definition 2.1 of a branching process, and we will exploit it extensively in the next chapter.

The random-walk formulation of branching processes allows us to study the probability of extinction when the family tree has at least a given size. In particular, the following theorem states that when the total progeny is large, then the probability of the branching process to become extinct tends to zero. Before stating and proving the theorem, we need to introduce some nomenclature.

Consider two functions, f and g, defined on some subset of the integers. We say that $f(n) = O(g(n))$ if and only if there exists a positive constant M and a number n_0 such that $|f(n)| < M|g(n)|$, $\forall n > n_0$.

Theorem 2.8. *Let $X_1, \ldots, X_n$ be the family-size random variables of the branching process with total progeny T. If the mean of the generic family-size random variable is less or equal to 1, that is $E(X_i) \leq 1$, then*

$$P(k \leq T < \infty) = O\left(e^{-kI(1)}\right), \tag{2.7}$$

where $I(1)$ is the rate function of X_i in 1.

Proof. Let S_i be the number of active vertices after the i-th exploration. As noted before the process becomes extinct when $S_i = 0$ for the first time. Thus

$$P(k \leq T < \infty) = \sum_{i=k}^{\infty} P(\inf\{i : S_i = 0\}) \leq \sum_{i=k}^{\infty} P(S_i = 0).$$

Because of equation (2.5), we also know that

$$\sum_{i=k}^{\infty} P(S_i = 0) = \sum_{i=k}^{\infty} P(X_1 + \cdots + X_i = i - 1) \leq \sum_{i=k}^{\infty} P(X_1 + \cdots + X_i \leq i).$$

Now, because of the Chernoff bound in equation (1.10) and after some easy calculations, we can notice that

$$\begin{aligned} P(k \leq T < \infty) &\leq \sum_{i=k}^{\infty} e^{-sI(1)} \\ &= \sum_{i=0}^{\infty} e^{-sI(1)} - \sum_{i=0}^{k-1} e^{-sI(1)} \\ &= \frac{1}{1 - e^{-I(1)}} - \frac{1 - e^{-kI(1)}}{1 - e^{-I(1)}} \\ &= \frac{e^{-kI(1)}}{1 - e^{-I(1)}} = O\left(e^{-kI(1)}\right). \end{aligned}$$

□

2.4 Binomial and Poisson branching processes

In this section we are going to introduce a result that we will use extensively in the next chapter. We will consider a binomial model for random graphs and, as said, we will compare the components to branching processes with a binomial offspring distribution. Moreover, we will use the Poisson approximation to relate a branching process with binomial offspring distribution to another process with Poisson offspring distribution. For this reason we introduce now the so-called Poisson branching process and then we will see the relationship between binomial and Poisson branching processes.

Poisson and binomial branching processes

We say that a branching process is a **Poisson branching process** with mean λ if the offspring distributions are Poisson random variables with mean λ. For this kind of branching processes we can compute all the characteristics introduced in the previous section. We start by computing the extinction probability.

Recalling that, if $X \sim P(\lambda)$, the probability generating function of X in s is $e^{\lambda(s-1)}$, we can compute the probability that a Poisson branching process with mean λ dies out, which can be found from the equation

$$s = \exp(\lambda(s-1)).$$

If $\lambda > 1$, the probability of extinction is $1 - \beta(\lambda)$, where $\beta = \beta(\lambda) \in (0,1)$ is the survival probability and β can be uniquely determined by the equation

$$\beta + \exp(-\beta\lambda) = 1, \tag{2.8}$$

since

$$1 - \beta = \exp(\lambda(1 - \beta - 1)) = \exp(-\beta\lambda).$$

We can also compute the total expected progeny of such a process, since, from equation (2.3), it is simply

$$\frac{1}{1-\lambda}.$$

It is also possible to know something more about the total progeny of a Poisson branching process, and in particular we can compute the probability mass function of the total progeny.

Theorem 2.9. *Let T be the total progeny of a Poisson branching process with mean $\lambda > 0$. Then*

$$P(T = n) = \frac{e^{-\lambda n}(\lambda n)^{n-1}}{n!}$$

Proof. We will exploit in this proof the random-walk representation of a branching process. We start by noting that, in the random-walk setting, the total progeny, as defined in (2.6), is simply the hitting time of the origin, as in (1.13).

For this reason, we can use Theorem 1.19 to deduce that[4]

$$\begin{aligned} P_1(T=n) &= \frac{1}{n}P_1(S_n=0) \\ &= \frac{1}{n}P_1(Y_1+\cdots Y_n=0) \\ &= \frac{1}{n}P_1(X_1+\cdots X_n=n-1), \end{aligned}$$

where the generic X_i is the family-size distribution (in this case a Poisson random variable with mean λ) and $Y_i = X_i - 1$. Since the X_i's are also independent, the sum is a Poisson random variable with mean $n\lambda$ and the result follows. □

We introduce also the binomial branching process. Intuitively, we say that a branching process is a *binomial branching process* with parameters (n, p) if the offspring distribution is binomial with parameters (n, p). We will not need any further consideration about this kind of process and, thus, we just proceed to the comparison of the two processes.

Comparison of the two processes

We define now some notation before stating next theorem: we say that $f(n) = o(g(n)))$ as $n \to \infty$, if $g(n) > 0$, $n \in \mathbb{N}$, and $\lim_{n\to\infty} |f(n)|/g(n) = 0$. This means, roughly speaking, that $g(n)$ goes to the limit faster than $f(n)$.

The theorem that we are going to introduce states that, when n gets bigger, the probability that the total progenies of a binomial and a Poisson branching process are different is almost surely zero.

Theorem 2.10. *Let $T^B_{n,p}$ be the total progeny of a binomial branching process with parameters (n, p) and let also T^P_λ be the total progeny of a Poisson branching process with mean λ, where $\lambda = np$. Then, for any $k \geq 1$*

$$P(T^B_{n,p} \geq k) = P(T^P_\lambda \geq k) + O(k/n), \tag{2.9}$$

which, alternatively, can also be written as

$$P(T^B_{n,p} \geq k) = P(T^P_\lambda \geq k) + o(1) \tag{2.10}$$

Proof. We start this proof by simply noting that, from the law of total probability,

$$P(T^B_{n,p} \geq k) = P(T^B_{n,p} \geq k, T^P_\lambda \geq k) + P(T^B_{n,p} \geq k, T^P_\lambda < k),$$

$$P(T^P_\lambda \geq k) = P(T^P_\lambda \geq k, T^B_{n,p} \geq k) + P(T^P_\lambda \geq k, T^B_{n,p} < k).$$

Thus,

$$\begin{aligned} P(T^B_{n,p} \geq k) - P(T^P_\lambda \geq k) &= P(T^B_{n,p} \geq k, T^P_\lambda < k) - P(T^P_\lambda \geq k, T^B_{n,p} < k) \\ &\leq P(T^B_{n,p} \geq k, T^P_\lambda < k) + P(T^P_\lambda \geq k, T^B_{n,p} < k). \end{aligned}$$

Taking the modulus on both sides, we have that

$$|P(T^B_{n,p} \geq k) - P(T^P_\lambda \geq k)| \leq P(T^B_{n,p} \geq k, T^P_\lambda < k) + P(T^P_\lambda \geq k, T^B_{n,p} < k).$$

[4]Since we are relating to Theorem 1.19, we are now stressing the fact that the Branching Process starts with one only individual, by putting a 1 as subscript. However, these probabilities are the same as in (2.4) and (2.5).

Now, let $X_1^B, X_2^B, \cdots \overset{i.i.d.}{\sim} B(n,p)$ and let also $X_1^P, X_2^P, \cdots \overset{i.i.d.}{\sim} P(\lambda)$. Note that, from equation (2.6), $T_{n,p}^B \geq k$ if, for any $t < k$, $X_1^B + \cdots + X_t^B \neq t-1$. Similarly, $T_\lambda^P \geq k$, if, for any $t < k$, $X_1^P + \cdots + X_t^P \neq t-1$.

Let's focus on $P(T_{n,p}^B \geq k, T_\lambda^P < k)$. Since we are considering the event in which the total progenies of the two branching processes assume a different value, there has to be an $s < k$ for which $X_s^B \neq X_s^P$. If there were not such a value, then both the processes would have different total progenies. We can bound the probability of $\{T_{n,p}^B \geq k, T_\lambda^P < k\}$, by considering any case that may lead to a different total progeny: this can be obtained by summing over all possible times $s < k$ in which the two processes are different for the first time, that is

$$P(T_{n,p}^B \geq k, T_\lambda^P < k) \leq \sum_{s=1}^{k-1} P(X_s^B \neq X_s^P, X_i^B = X_i^P,\ \forall i < s, T_{n,p}^B \geq k).$$

Now note that the event $\{T_{n,p}^B \geq k, X_i^B = X_i^P,\ \forall i < s\}$ implies that the total progeny of the Poisson branching process is bigger or equal than s, that is $T_\lambda^P \geq s$. Moreover, this event, $\{T_\lambda^P \geq s\}$, is independent of the event $\{X_s^B \neq X_s^P\}$, since the first one is only determined by the random variables $X_1^B, \ldots, X_{s-1}^B$. Thus

$$\begin{aligned}
P(T_{n,p}^B \geq k, T_\lambda^P < k) &\leq \sum_{s=1}^{k-1} P(X_s^B \neq X_s^P, X_i^B = X_i^P,\ \forall i < s, T_{n,p}^B \geq k) \\
&= \sum_{s=1}^{k-1} P(T_\lambda^P \geq s, X_s^B \neq X_s^P, T_{n,p}^B \geq k) \\
&\leq \sum_{s=1}^{k-1} P(T_\lambda^P \geq s, X_s^B \neq X_s^P) \\
&= \sum_{s=1}^{k-1} P(T_\lambda^P \geq s) P(X_s^B \neq X_s^P) \\
&\leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} P(T_\lambda^P \geq s).
\end{aligned}$$

The last inequality follows from Theorem 1.11, that is, the Poisson approximation. Now, by the same reasoning, we can say that

$$P(T_\lambda^P \geq k, T_{n,p}^B < k) \leq \sum_{s=1}^{k-1} P(X_s^B \neq X_s^P, X_i^B = X_i^P\ , \forall i < s, T_\lambda^P \geq k).$$

In this second case, the reasoning is much more easier. Indeed, by simply noting that, for any $s \leq k$, $P(T_\lambda^P \geq k) \leq P(T_\lambda^P \geq s)$, and that $\{X_s^B \neq X_s^P, T_\lambda^P \geq k\} \subseteq$

$\{X_s^B \neq X_s^P, X_i^B = X_i^P \ , \forall i < s, T_\lambda^P \geq k\}$, it follows that

$$\begin{aligned} P(T_\lambda^P \geq k, T_{n,p}^B < k) &\leq \sum_{s=1}^{k-1} P(X_s^B \neq X_s^P, X_i^B = X_i^P \ , \forall i < s, T_\lambda^P \geq k) \\ &\leq \sum_{s=1}^{k-1} P(X_s^B \neq X_s^P, X_i^B = X_i^P, \ \forall i < s, T_\lambda^P \geq s) \\ &\leq \sum_{s=1}^{k-1} P(X_s^B \neq X_s^P, T_\lambda^P \geq s) \\ &\leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} P(T_\lambda^P \geq s). \end{aligned}$$

Finally,

$$\begin{aligned} |P(T_{n,p}^B \geq k) - P(T_\lambda^P \geq k)| &\leq P(T_{n,p}^B \geq k, T_\lambda^P < k) + P(T_\lambda^P \geq k, T_{n,p}^B < k) \\ &\leq \frac{2\lambda^2}{n} \sum_{s=1}^{k-1} P(T_\lambda^P \geq s) \\ &\leq \frac{2k\lambda^2}{n} = O(k/n) = o(1), \end{aligned}$$

from which the result easily follows. □

Chapter 3

The phase transition

After the first two chapters, in which we introduced all the tools necessary for the analysis of random graphs, we are now ready in this third chapter to work on the main theme of the dissertation: the phase transition of the largest component size of random graphs. We will always consider in this chapter a binomial random graph $G(n, c/n)$, where $n \in \mathbb{N}$ and $c \in (0, n)$.[1] In the first section we will describe how the components of a random graph are explored, while the second one will consist of a gentle discussion about the phase transition, followed by a simple simulation. At the end of the section, however, the result of the phase transition will be formalized and we will state the main theorem of the dissertation. In the third section we will relate the size of the largest component to another quantity: this will become handy during the proof of Theorem 3.1, which will be developed in the fourth and the fifth sections. Finally, in the last section we will show which is the size of the largest component in the critical case. The whole chapter is mainly based on [5].

3.1 The exploration of the components

Since the main focus of this section is on the components of a random graph, we have to introduce some notations about them. First of all, from now on, we will consider the vertex set to be labeled with a subset of the integer set: that is, vertices are $1, 2, 3, \ldots$ and so on. We will also denote the component of vertex v by $C(v)$, and its size by $|C(v)|$. The size of the largest component of a random graph, which is the object of main interest, is defined as [2]

$$|C_{\max}| = \max_{v \in \{1,\ldots,n\}} |C(v)|. \tag{3.1}$$

We can now describe the procedure of exploration. We have a graph with n vertices and we start by looking at the vertex with label 1. If this vertex has some incident edges, it means that $|C(1)| > 1$ and, thus, vertex 1 has some

[1] It is not mandatory to exclude the cases in which $c = 0$ or $c = n$. However, they correspond, respectively, to the degenerate cases of a graph with no edges and of a complete graph.

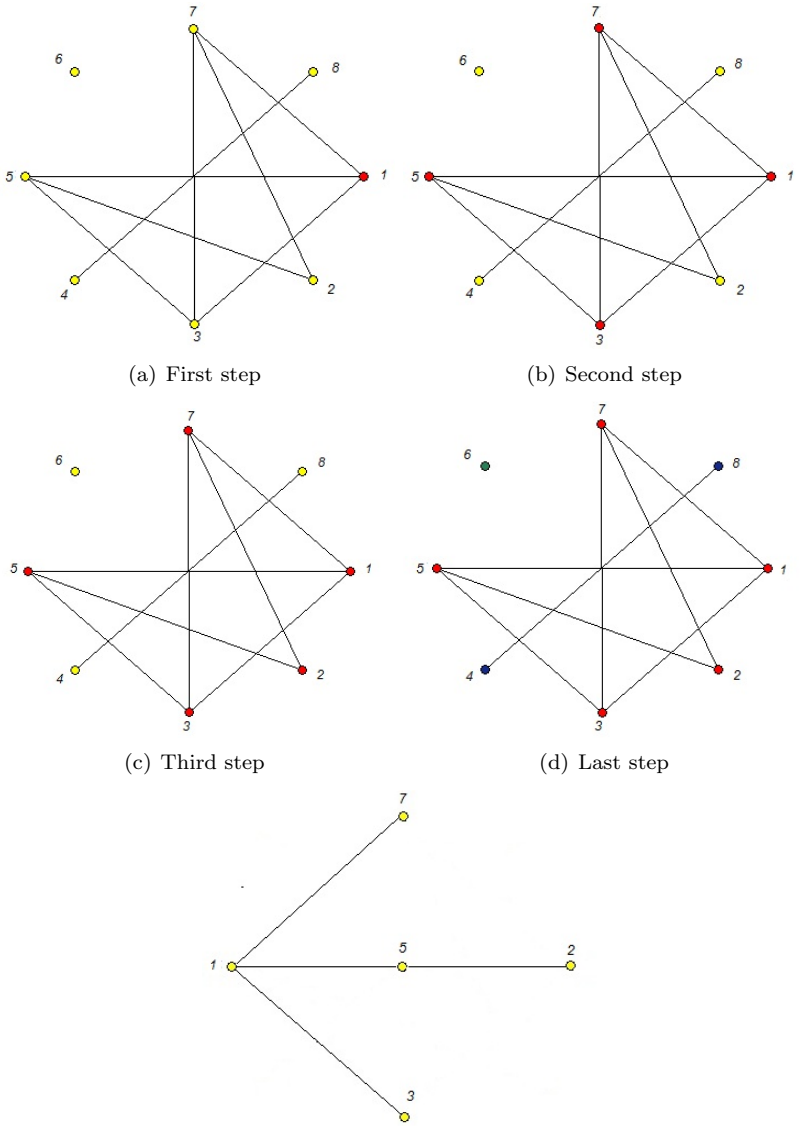
[2] Note that equation 3.1 does not necessarily identify uniquely the largest component, since there may be more components with the biggest number of vertices. It is possible to make the definition unique by requiring C_{max} to be the cluster containing the vertex with the smallest label among the components with maximal size.

neighbors. We can think of the neighbors of vertex 1 as like the first generation of a branching process with root vertex 1. So, after that, we choose one of the neighbors of vertex 1 and we look at its neighbors, without considering the neighbors of 1 and 1 itself. We can think of the found vertices at this step like a part of the second generation of the branching process with root 1. At this point, we can use one of the two algorithms introduced in section 2.3, in order to pick the next vertices to look at, between those we have already seen. The process can go on in this fashion until there are vertices that we have already seen, but of which we haven't look at their neighbors. Once we have looked at all the vertices that are in some generation of the branching process with root 1, it means that we have fully explored a component of the random graph. At this point, we have to pick the vertex with the lowest label between the ones available and start a new exploration, only considering vertices not in $C(1)$, and so on. In order to understand the process, we do the exploration on the graph of Figure (3.1). Figure (a) shows a random graph, in which vertex 1 is chosen, in order to look at its neighbors. This vertex is colored in red: this color will represent the vertex that are seen in the first component, while yellow corresponds to vertices that are not seen yet. In figure (b) is shown the result of the first exploration, that is, the neighbors of vertex 1 (vertices 3, 5 and 7). At this point, vertex 3 is considered: however, no yellow vertices are neighbors of this vertex and so nothing happens. Then we look at vertex 5 and we can notice that vertex 2 is yellow and it is a neighbor of 5: for this reason in figure (c) it becomes red. Vertices 7 and 2 do not have any yellow neighbor and so the exploration of $C(1)$ ends. Figure (e) shows the tree that is related to the exploration of the first component of the random graph: the root vertex is 1, the first generation consists of vertices 3, 5 and 7, which are the neighbors of 1, while in the second one there is only vertex 2. It does not matter if in the random graph there are some edges between the vertices that belongs to the first generation: indeed, when we look at one of them, the others are already red and, for this reason, it is not significant the presence or less of an edge joining them. After completing the exploration of the first component, we pick the yellow vertex with the lowest index, which, in this case, is 4, and we start the exploration of its component (denoted by the color blue). Figure (d) shows the final result of the exploration of the components with the blue vertices 4 and 8 in the second component and vertex 6 in the third one with a green color.

The one just described is the general procedure of exploration of the components. However, we can relate it to the random-walk perspective of branching processes and use the same nomenclature active-neutral-inactive for the vertices.[3] We also specialize now to the binomial model of random graphs, since the description above is true for any kind of graph.

Consider a graph G in which all the vertices are considered neutral, except for vertex 1, which is the only one active: so we are considering $S_0 = 1$, where S represents the number of active vertices, as in section 2.3. Then, we look at its neighbors: the number of vertices that are joined through an edge to 1 is the result of a binomial random variable with parameters $(n-1, c/n)$, since we have $n-1$ possible neighbors (neutral vertices), each of them having a probability c/n

[3] In [12] the exploration of the components works in the same way but a different nomenclature is used. In particular the authors denote with a name also the edges of the graph, depending on the kind of vertices that they join. However, this kind of notation is a little bit harder and redundant: for this reason the notation of [5] is preferred.



(a) First step

(b) Second step

(c) Third step

(d) Last step

(e) Tree of the exploration of $C(1)$

Figure 3.1: Steps of the exploration of the components.

of being joined to 1. We denote with X_1 this binomial random variable. In the example of Figure (3.1), $X_1 \sim B(7, c/n)$, which takes value 3 (vertex 1 has three neighbors)[4]. Moreover, the neighbors of 1 become active vertices, while vertex 1 becomes inactive: thus, in the example $S_1 = 3$, where $S_1 = X_1 - S_0$ represents the number of active vertices after the first exploration. At this point, we have to pick one of the active vertices and see if there are some of its neighbors between the neutral vertices. In this case, the result comes from the binomial random variable X_2 with parameters $(n - X_1 - 1, c/n)$. Again, the number of active

[4]We are not interested in the value of the second parameter c/n at the moment.

vertices is updated and the chosen vertex becomes inactive. In the example, the first parameter of X_2 is 4 and, if vertex 3 is chosen, it has not any neighbor in the neutral set: thus $S_2 = 2$. Clearly, we can represent this procedures as a random walk, by simply noting that

$$S_0 = 1, \qquad S_i = S_{i-1} + X_i - 1, \tag{3.2}$$

where S_{i-1} is the set of active vertices before the i-th exploration and X_i represents the result of the i-th exploration: it is a binomial random variable with parameters $(n - (X_1 + \cdots X_{i-1}) - 1), c/n)$. We can also write the first parameter of X_i as $n - S_{i-1} - (i-1)$, since $X_1 + \cdots X_{i-1} = S_{i-1} - (i-1) + 1$, and note that this simply corresponds to the size of the neutral set.

The process of exploration of the first component continues in this way, until there are active vertices to explore. So $|C(1)| = \inf\{t : S_t = 0\}$, which is equal to the T defined in equation (2.6): thus the size of the first component corresponds to the total progeny of the related branching process. Once the first component is fully explored, we have to pick one vertex in the neutral set, switch its state to active and continue with the exploration of the other components.

However, when a component is fully explored, there is a little issue with the number of active vertices: for this reason we define

$$S_i^1 = S_i - |\{\text{components fully explored after the i-th exploration}\}|$$

which is simply the number of active vertices with a correction term that corresponds to the number of components fully explored. This new quantity is necessary to maintain the random walk construction of the exploration, after the first component is fully explored: indeed, while the first component is explored, $S_i^1 = S_i$ and, thus, we can just replace S_i with S_i^1 in (3.2). However, this is no longer true for the other components. Consider S_i: when $S_i = 0$, a component is fully explored and there are no more active vertices to look at. However, we need a new active vertex to start a new component exploration, that is, S_i should be equal to 1. This problem is circumvented by considering S^1: indeed, when $S_i = 0$ for the first time, automatically the neutral vertex with lowest index switch its state and the number of components fully explored becomes one. In this way, we have an active vertex to look at, the value of S_i^1 remains 0, as it should be, and we can extend the random walk construction to the exploration of all the components. Note that now the second component will be fully explored when $S_i^1 = -1$ for the first time, while the third one is obtained when $S_i^1 = -2$ and so on. Let also k be the lowest-labeled neutral vertex after the first component is fully explored, then the size of the second component is

$$|C(k)| = \inf\{t : S_t^1 = -1\} - |C(1)|.$$

In Figure 3.2 we can see a graph that represents the process of exploration: dots corresponds to the steps in which end a component's exploration.

We can also summarize the various steps of the exploration process in a table, as shown in Table 3.1. The first column represents the number of the exploration (in the first line there is a zero to consider the initial state), while in the second one we can find the label of the vertex chosen for the exploration. The third one consists of the number of active vertices after the i-th exploration: in two cases we can note the value 0/1, which is caused by the problem that arises when a

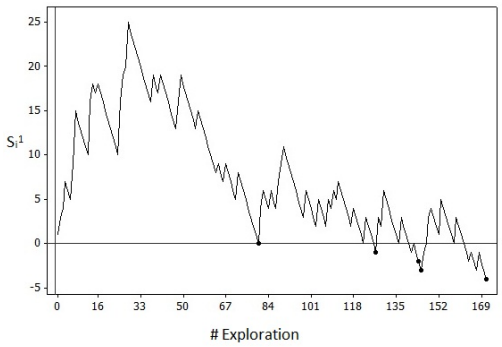


Figure 3.2: Representation of a graph exploration.

	Vertex	$\|\mathbf{S_i}\|$	$\|\mathbf{N_i}\|$	$\|\mathbf{I_i}\|$	**Num. Comp.**	$\|\mathbf{S_i^1}\|$
0		1	7	0	0	1
1	1	3	4	1	0	3
2	3	2	4	2	0	2
3	5	2	3	3	0	2
4	2	1	3	4	0	1
5	7	0/1	2	5	1	0
6	4	1	1	6	1	0
7	8	0/1	0	7	2	-1
8	6	0	0	8	3	-1

Table 3.1: Summary of the state of the vertices during the exploration

component is fully explored. The fourth and the fifth ones represent, respectively, the number of neutral and inactive vertices after the i-th exploration, while the sixth column contains the number of components fully explored. Finally, the last column consists of the number of the so-called corrected-active vertices.

Before finishing this first section, there is an observation to be made. The random variables that represent the family size in a branching process are a sequence of i.i.d. random variables. For this reason, generations far from the root have the same probability of generations close to the root of having a certain number of children. This does not happen in the exploration of a random graph. First of all, the binomial random variables are not independent and are not identically distributed: for example, the first parameter of X_2 is $n - X_1 - 1$, which obviously depends on the value of X_1. Moreover, the number of neutral vertices decreases as the number of exploration increases: for this reason, it is less probable that a vertex, which is explored quite late, has many neighbors. This effect is also know as **depletion of points**.

3.2 Size of the largest component

In the previous chapter, and in particular in Section 2.2, we noticed a phase transition for the mean value of the generation size of branching processes. There

we stated that the importance of branching processes in this context was all about this shared behavior with the largest component size of random graphs. We are now ready to introduce this characteristic behavior for random graphs too. Consider a binomial random graph with parameter $(n, c/n)$: depending on the value of c, the size of the largest component varies with a phase transition behavior. If $c < 1$ (**subcritical phase**), the size of the largest component is of the order of $\ln(n)$. This means that there are many components of small dimension: for this reason the random graph is defined **sparse**. On the contrary, if $c > 1$ (**supercritical phase**), the size of the largest component is of the order of n: there is a very big component, which contains a positive proportion of the vertices. We will also prove that, in the supercritical phase, there are no other components with sizes of order between $\ln(n)$ and n: thus, the largest component is much bigger than the others. For this reason, we call the largest component the **giant component,** and, moreover, we also refer to this phase transition as the *emergence of the giant component*.

Note that, the mean number of neighbors of a vertex in the graph considered is $(n-1)\frac{c}{n}$, which asymptotically is equal to c. This means that, if the mean number of edges incident to a vertex is less than one, the graph is in the subcritical phase, while if it is bigger than one, the graph is in the supercritical phase. This last consideration allows us to link again the exploration of a random graph to branching processes. Indeed, if the expected number of neighbors for a vertex in a random graph is less than one, this is also true for the related branching process. However, this means that, on average, every individual of the branching process will generate less than one offspring, causing the extinction of the process. On the contrary, when we are in the supercritical phase, every individual of the related branching process will have, on average, more than one offspring: in this situation we know that there is a certain probability that the process does not die out. Because of the depletion of points effect, this cannot happen for a random graphs exploration and its related branching process. However, we can deduce that, in this setting, it is probable to find a related branching process with many generations, which corresponds to a large component.

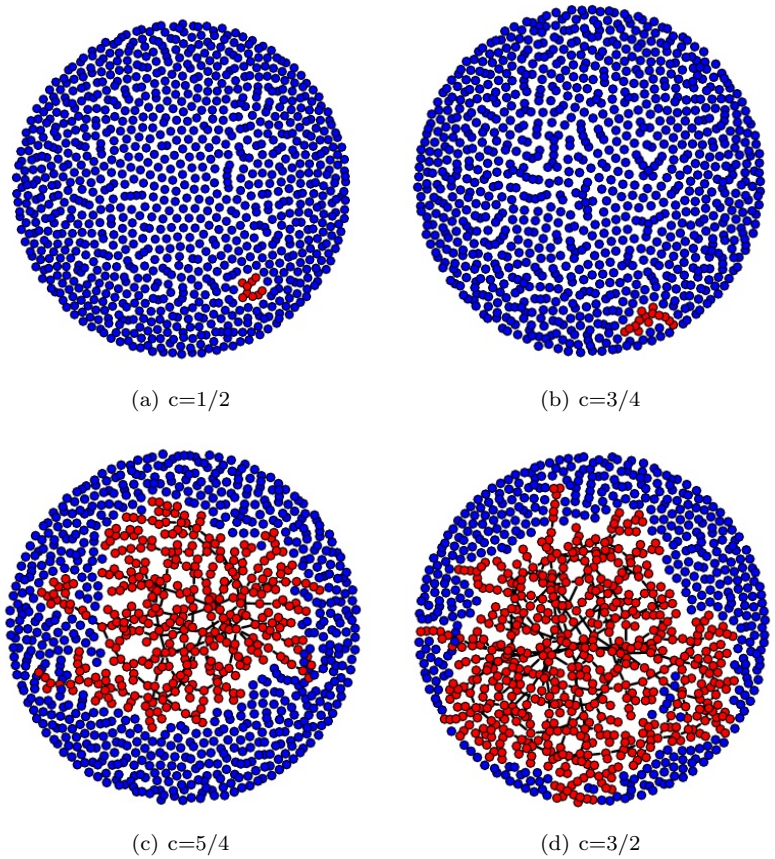
We can think of the random graph process in order to understand in a better way the behavior of the size of the largest component. Recall that, the random graph process considers a sequence of graphs, the first of them with no edge, and then it adds new edges, one at a time. We can easily see that this process considers a uniform model of random graphs, since the number of edges is fixed at every step. Note also that the expected number of edges in a binomial model is $\binom{n}{2}\frac{c}{n}$: so we can relate the two models by considering $c = \frac{2M}{n-1} \approx \frac{2M}{n}$, where M is the number of edges of the uniform model. It is important to underline that in this observation we are considering the mean number of edges of the binomial random graph. In the random graph process the value of M increases of a unity at every step: so until the value of M is such that $c < 1$, we are in the subcritical phase. In this phase, at every step one edge is added: however, the number of edges is small respect to the number of vertices, and, so, it is hard that a high number of edges joins vertices to the same component. It is more likely that the new edge joins vertices that do not belong to any `"larger"` component. However, as the number of edges increases, it is possible to create small assemblages of vertices, each of them of small dimension. When the number of edges is large enough to make c close to one, the graph has many components of size close to $\ln(n)$ and at this point it is more likely that an edge does not

joins single vertices, but vertices that belong to different larger components. This in particular happens when $c = 1$ and we are in the **critical phase**. We will show in the last section that, in this case, the size of the largest component is of the order $n^{2/3}$. So, as said, in this phase it is easier that two components of order $\ln(n)$ are joined, with the result that, when we reach the supercritical phase, many of these components are now a single giant component.

We now simulate some random graphs in order to show the phase transition behavior. Every graph will consists of 1000 vertices and we will consider four different edge probabilities.

	1st	**2nd**	**3rd**	**4th**
1/2	9	6	6	5
3/4	14	12	10	9
5/4	359	44	11	8
3/2	562	20	10	10

Table 3.2: Size of largest components.



(a) c=1/2 (b) c=3/4

(c) c=5/4 (d) c=3/2

Figure 3.3: Simulation of random graphs with different edge probabilities

In Table 3.2 are summarized the results of the simulation. In the subcritical phase, which corresponds to the first two lines, we can notice that the largest component size (first column) is quite small and is comparable to the value $\ln(1000) \approx 7$. Moreover, the values of the sizes of the other components (other columns) is close to the largest one in both cases. This fact, as expected, does not happen in the supercritical phase (last two rows), in which, clearly, there is a component much larger than the others. In Figure 3.3, is proposed a visual representation of the graphs simulations. Red dots belongs to the largest component: in cases (a) and (b) the component is really small and does not distinguish itself from the others. In the two other cases, the largest component is easy to identify and occupies a great portion of the graph. It is important to underline that this simulation is not the result of a random graph process, but simply the realizations of some binomial random graphs with different edge probabilities.

We are now ready to state the Theorem about the phase transition of the largest component size.

Theorem 3.1. *Let G be a binomial random graph with parameters $(n, c/n)$ and let also $I_P = I_P(1) = c - 1 - log(c)$ be the rate deviation function of a Poisson random variable with mean c.*
If $c < 1$ and $a > I_c^{-1}$, then

$$\lim_{n\to\infty} P(|C_{max}| \geq a \ln(n)) = 0, \tag{3.3}$$

while if $c < 1$ and $b < I_c^{-1}$, we have that

$$\lim_{n\to\infty} P(|C_{max}| < b \ln(n)) = 0. \tag{3.4}$$

If $c > 1$ and β_c is the survival probability of a Poisson branching process with mean c, then for every $v \in (\frac{1}{2}, 1)$

$$\lim_{n\to\infty} P(|\,|C_{max}| - \beta_c n| > n^v) = 0 \tag{3.5}$$

equations (3.3) and (3.4) concern the subcritical regime and, together, they imply that, with a probability that tends to one, the value of the size of the largest component is in the interval $[a, b]$, which includes the value $I_c^{-1} \ln(n)$. On the contrary, in the supercritical phase, with a probability that tends to one, the number of vertices in the largest component belongs to an interval which includes $\beta_c n$, with amplitude $2|n^v - \beta_c n|$.

3.3 Another representation of the largest component size

In the next two sections we will prove the fact that the largest component is of a certain size, depending on the value of c. This proof will be based largely on a different representation of the largest component, that we now introduce. Let

$$|C^{\geq k}| = \sum_{v=1}^{n} \mathbf{1}_{\{|C(v)|\geq k\}} \tag{3.6}$$

denote the number of vertices that belongs to components of size at least k, where $\mathbf{1}$ is the indicator function, defined as

$$\mathbf{1}_{\{|C(v)|\geq k\}} = \begin{cases} 1 & \text{if } |C(v)| \geq k \\ 0 & \text{if } |C(v)| < k. \end{cases}$$

The following theorem shows the relationship between this new quantity and the size of the largest component.

Theorem 3.2. *Let $|C^{\geq k}|$ be defined as in equation (3.6) and let $k < n$, $k, n \in \mathbb{N}$, then*

$$|C_{max}| = \max\{k : |C^{\geq k}| \geq k\}. \tag{3.7}$$

Moreover

$$\{|C_{max}| \geq k\} = \{|C^{\geq k}| \geq k\}. \tag{3.8}$$

Proof. If $|C_{max}| < k$, then $|C(v)| < k$ for every vertex v and thus $|C^{\geq k}| = 0$. On the contrary, if $|C_{max}| \geq k$, then $|C(v)| \geq k$ for at least k vertices and thus $|C^{\geq k}| \geq k$. For this reason, while $k \leq |C_{max}|$, the value of $|C^{\geq k}| \geq k$, but as soon as the value of k exceed the size of the largest component $|C^{\geq k}| = 0$: thus equation (3.7) follows.

To prove (3.8), we have to show that if $|C^{\geq k}| \geq k \Longrightarrow |C_{max}| \geq k$, since we have already proved the other direction. However, by definition, the presence of more than k vertices in components of size at least k implies that the size of the largest component is at least k. □

Now, we compute the first moment of the quantity $|C^{\geq k}|$ and two bounds of its second moment.

Theorem 3.3. *Let $|C^{\geq k}|$ be defined as in equation (3.6) and c/n be the edge probability of a binomial random graph with n vertices. Then, for $k, n \in \mathbb{N}$ with $k < n$,*

$$E(|C^{\geq k}|) = nP(|C(v)| \geq k), \tag{3.9}$$

$$Var(|C^{\geq k}|) \leq nE\left(|C(v)|\mathbf{1}_{\{|C(v)|\geq k\}}\right), \tag{3.10}$$

$$Var(|C^{\geq k}|) \leq (ck+1)nE\left(|C(v)|\mathbf{1}_{\{|C(v)|<k\}}\right). \tag{3.11}$$

Proof. For the expectation, note that

$$\begin{aligned} E(|C^{\geq k}|) &= E\left(\sum_{v=1}^{n} \mathbf{1}_{\{|C(v)|\geq k\}}\right) \\ &= \sum_{v=1}^{n} E\left(\mathbf{1}_{\{|C(v)|\geq k\}}\right) \\ &= \sum_{v=1}^{n} P\left(|C(v)| \geq k\right) \\ &= nP\left(|C(v)| \geq k\right). \end{aligned}$$

Now, for the proof of equations (3.10), let's start by considering the definition of the variance:

$$\begin{aligned} Var(|C^{\geq k}|) &= E\left(\left(\sum_{v=1}^{n}\mathbf{1}_{\{|C(v)|\geq k\}}\right)^2\right) - E\left(\sum_{v=1}^{n}\mathbf{1}_{\{|C(v)|\geq k\}}\right)^2 \\ &= E\left(\left(\sum_{v=1}^{n}\mathbf{1}_{\{|C(v)|\geq k\}}\right)\left(\sum_{j=1}^{n}\mathbf{1}_{\{|C(j)|\geq k\}}\right)\right) \\ &\quad - \left(\sum_{v=1}^{n}P(|C(v)|\geq k)\right)^2 \\ &= E\left(\sum_{v,j=1}^{n}\mathbf{1}_{\{|C(v)|\geq k,|C(j)|\geq k\}}\right) \\ &\quad - \left(\sum_{v=1}^{n}P(|C(v)|\geq k)\right)\left(\sum_{j=1}^{n}P(|C(j)|\geq k)\right) \\ &= \sum_{v,j=1}^{n}\left[P(|C(v)|\geq k,|C(j)|\geq k) - P(|C(v)|\geq k)P(|C(j)|\geq k)\right]. \end{aligned} \tag{3.12}$$

Now, by the law of total probability we can note that

$$\begin{aligned} P(|C(v)|\geq k,|C(j)|\geq k) &= P(|C(v)|\geq k,\{v,j \text{ connected}\}) \\ &+ P(|C(v)|\geq k,|C(j)|\geq k,\{v,j \text{ not connected}\}), \end{aligned}$$

from which we can rewrite the variance as

$$\begin{aligned} Var(|C^{\geq k}|) = \sum_{v,j=1}^{n} &[P(|C(v)|\geq k,\{v,j \text{ connected}\}) \\ &+ P(|C(v)|\geq k,|C(j)|\geq k,\{v,j \text{ not connected}\}) \\ &- P(|C(v)|\geq k)P(|C(j)|\geq k)] \\ &\leq \sum_{v,j=1}^{n} P(|C(v)|\geq k,\{v,j \text{ connected}\}), \end{aligned}$$

since

$$\begin{aligned} &P(|C(v)|\geq k,|C(j)|\geq k,\{v,j \text{ not connected}\}) \\ &\leq P(|C(v)|\geq k,|C(j)|\geq k) \\ &\leq P(|C(v)|\geq k)P(|C(j)|\geq k). \end{aligned}$$

Finally, we can deduce that, using again the properties of the indicator function,

$$\begin{aligned} Var(|C^{\geq k}|) &\leq \sum_{v,j=1}^{n} P(|C(v)| \geq k, \{v, j \text{ connected}\}) \\ &= \sum_{v,j=1}^{n} E\left(\mathbf{1}_{\{|C(v)\geq k|\}}\mathbf{1}_{\{j\in C(v)\}}\right) \\ &= \sum_{v=1}^{n} E\left(\mathbf{1}_{\{|C(v)|\geq k\}}\sum_{j=1}^{n}\mathbf{1}_{\{j\in C(v)\}}\right) \\ &= \sum_{v=1}^{n} E\left(|C(v)|\mathbf{1}_{\{|C(v)|\geq k\}}\right) \\ &= nE\left(|C(v)|\mathbf{1}_{\{|C(v)|\geq k\}}\right) \end{aligned}$$

Consider now equation (3.11) and define

$$|C^{<k}| = \sum_{v=1}^{n} \mathbf{1}_{\{|C(v)<k|\}}.$$

Then, clearly $|C^{<k}| = n - |C^{\geq k}|$ and thus $Var(|C^{<k}|) = Var(|C^{\geq k}|)$. For this reason, we will prove that

$$Var(|C^{<k}|) \leq (ck+1)nE\left(|C(v)|\mathbf{1}_{\{|C(v)|<k\}}\right).$$

By the same reasoning developed to obtain (3.12), we can derive that

$$Var(|C^{<k}|) \sum_{v,j=1}^{n} \left[P(|C(v)| < k, |C(j)| < k) - P(|C(v)| < k)P(|C(j)| < k)\right],$$

and again using the law of total probability, we can write

$$\begin{aligned} Var(|C^{<k}|) &= \sum_{v,j=1}^{n} [P(|C(v)| < k, |C(j)| < k, \{v, j \text{ not connected}\}) \\ &\quad - P(|C(v)| < k)P(|C(j)| < k)] \\ &\quad + \sum_{v,j=1}^{n} [P(|C(v)| < k, |C(j)| < k, \{v, j \text{ connected}\})] \end{aligned} \tag{3.13}$$

Now note that, by the same reasoning as the one developed to derive equation (3.10), we can write the second sum of the right-hand side of (3.13) as

$$\sum_{v,j=1}^{n} [P(|C(v)| < k, |C(j)| < k, \{v, j \text{ connected}\})] = nE\left(|C(v)|\mathbf{1}_{\{|C(v)|<k\}}\right)$$

On the contrary, for the first sum of the right-hand side of (3.13), note that

$$\begin{aligned} &P(|C(v)| < k, |C(j)| < k, \{v, j \text{ not connected}\}) = \\ &\qquad \sum_{i=1}^{k-1} P(|C(v)| = i, |C(j)| < k, \{v, j \text{ not connected}\}). \end{aligned}$$

Now, consider $P(|C(v)| = i, |C(j)| < k, \{v, j \text{ not connected}\})$: by the factorization theorem, we can write this probability as

$$\begin{aligned} P(|C(v)| = i, |C(j)| < k, \{v, j \text{ not connected}\}) = \\ P(|C(v)| = i) P(\{v, j \text{ not connected}\} \mid |C(v)| = i) \\ P(|C(j)| < k \mid \{v, j \text{ not connected}\}, |C(v)| = i). \end{aligned}$$

Moreover,

$$\begin{aligned} P(|C(v)| = i, |C(j)| < k, \{v, j \text{ not connected}\}) \leq \\ P(|C(v)| = i) P(|C(j)| < k \mid \{v, j \text{ not connected}\}, |C(v)| = i). \end{aligned} \quad (3.14)$$

The second probability on the right-hand side of (3.14) corresponds to the probability that the size of the component containing vertex j is less than k in a random graph with $n - i$ vertices and edge probability c/n. We denote this event by $P(|C_{n-i}(j)| < k)$ and thus

$$\begin{aligned} P(|C(j)| < k \mid \{v, j \text{ not connected}\}, |C(v)| = i) \\ = P(|C_{n-i}(j)| < k) \\ = P(|C(j)| < k) + P(|C_{n-i}(j)| < k) - P(|C(j)| < k). \end{aligned}$$

We can now construct a coupling between these two random graphs. Take the random graph with $n - i$ vertices and add the vertices $\{n - i + 1, \ldots, n\}$: then we will draw an edge between vertices s and t with probability c/n, where $s \in \{n - i + 1, \ldots, n\}$ and $t \in \{1, \ldots, n\}$. This is a coupling and, in particular, in this setting

$$P(|C_{n-i}(j)| < k) - P(|C(j)| < k) = P(|C_{n-i}(j)| < k, |C(j)| \geq k). \quad (3.15)$$

In order to prove equation (3.15), call $A = \{|C_{n-i}(j)| < k\}$ and $B = \{|C(j)| \geq k\}$: then $B^c = \{|C(j)| < k\}$ and $P(|C_{n-i}(j)| < k, |C(j)| \geq k) = P(A \cap B)$. By the basic rules of probability, we can derive that

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

However, $P(A \cup B) = 1$ since it contains any possible size of the component of the random graph; thus

$$P(A \cap B) = P(A) + P(B) - 1 = P(A) - P(B^c),$$

as required.

Now consider the event $\{|C_{n-i}(j)| < k, |C(j)| \geq k\}$: if $|C_{n-i}(j)| < k$, then $|C(j)| \geq k$ if and only if at least one of the vertices $\{n-i+1, \ldots, n\}$ is connected to one of the vertices belonging to the component of j. This happens, at most, with probability ikc/n, and thus

$$P(|C_{n-i}(j)| < k) - P(|C(j)| < k) \leq \frac{ikc}{n}.$$

Therefore,

$$\begin{aligned} Var(|C^{<k}|) &\leq nE\left(|C(v)|\mathbf{1}_{\{|C(v)|<k\}}\right) + \Bigg[\sum_{v,j=1}^{n}\sum_{i=1}^{k-1} P(|C(v)|=i) \\ &\quad \left(\frac{ikc}{n} + P(|C(j)|<k)\right) - P(|C(v)|<k)P(|C(j)|<k)\Bigg] \\ &= nE\left(|C(v)|\mathbf{1}_{\{|C(v)|<k\}}\right) + \sum_{v,j=1}^{n}\sum_{i=1}^{k-1}\frac{ikc}{n}P(|C(v)|=i) \\ &= nE\left(|C(v)|\mathbf{1}_{\{|C(v)|<k\}}\right) + \frac{ck}{n}\sum_{v,j=1}^{n} E\left(|C(v)|\mathbf{1}_{\{|C(v)|<k\}}\right) \\ &= (ck+1)nE\left(|C(v)|\mathbf{1}_{\{|C(v)|<k\}}\right). \end{aligned}$$

□

3.4 The subcritical regime

In this section, we are going to prove equations (3.3) and (3.4) of Theorem 3.1.

Proof of equation (3.3)

In order to prove the first equation, we will make the following steps:

1. We will show that the size of a component of a binomial random graph with parameters $(n, c/n)$ is dominated by the the total progeny of binomial branching process with the same parameters;
2. The total progeny of the binomial branching process will be bounded, using the Chernoff bound;
3. We will exploit the relationship with $|C^{\geq k}|$ and Markov inequality to compute another bound for $P(|C_{max}| \geq a\ln(n))$, which at the limit tends to zero.

For the first point, let N_i denote the number of neutral vertices in the random graph after i explorations, and let $X_i \sim B(N_{i-1}, c/n)$ be the number of vertices that become active after the i-th exploration. If we let $Y_i \sim B(n - N_{i-1}, c/n)$, then

$$X_i^{bp} = X_i + Y_i$$

dominates X_i, as shown in Example 1.15, since $X_i^{bp} \sim B(n, c/n)$. Moreover, the total progeny of a binomial branching process with parameters $(n, c/n)$, denoted by $T^B_{n,c/n}$, can be defined as

$$\inf\{i : S_i^{bp} = X_1^{bp} + \cdots X_i^{bp} - (i-1) = 0\},$$

where $X_1^{bp}, \ldots, X_i^{bp} \overset{i.i.d.}{\sim} B(n, c/n)$. If we let S_i be the number of active vertices in the random graph after the i-th exploration, with the exploration process

starting from 1, then[5]

$$P(|C(1)| > k) = P(S_i > 0, \forall i \leq k).$$

However, since X_i is dominated by X_i^{bp}, we have that

$$P(S_i > 0, \forall i \leq k) \leq P(S_i^{bp} > 0, \forall i \leq k) = P(T > k).$$

Finally, since the process of exploration can start from any vertex, and not necessary from the one labeled 1, the above argument is true for any vertex, so that

$$P(|C(v)| > k) \leq P(T_{n,c/n}^B > k). \tag{3.16}$$

Now, we can start proving point number two, by simply noting that

$$\begin{aligned} P(T_{n,c/n}^B > k) &= P(S_i^{bp} > 0, \forall i \leq k) \\ &= P(X_1^{bp} > 0) \cdots P(X_1^{bp} + \cdots X_k^{bp} > k-1) \\ &\leq P(X_1^{bp} + \cdots X_k^{bp} \geq k) = P(S_k^{bp} \geq k). \end{aligned}$$

Since $S_k^{bp} \sim B(nk, c/n)$ and using the Chernoff bound (Theorem 1.7), it follows that

$$P(X_1^{bp} + \cdots X_k^{bp} \geq k) \leq e^{-kI_B(1)},$$

where I_B is the rate function of a Bernoulli random variable with mean c/n. By Proposition 1.6, in which we related the rate function of binomial and Poisson random variables, we have that

$$P(X_1^{bp} + \cdots X_k^{bp} \geq k) \leq e^{-kI_P},$$

with $I_P = c - 1 - \ln(c)$. If we link the first two points, we can deduce that

$$P(|C(v)| > k) \leq e^{-kI_P}. \tag{3.17}$$

We can now consider the third point of the proof: because of Theorem 3.2, we have that, for $a > I_P^{-1}$,

$$P(|C_{max}| \geq a\ln(n)) = P\left(|C^{\geq a\ln(n)}| \geq a\ln(n)\right). \tag{3.18}$$

Since $I_P > 0$, it is true that for n large enough $a\ln(n) \geq 1$, so that using Markov inequality and equation (3.9)

$$\begin{aligned} P\left(|C^{\geq a\ln(n)}| \geq a\ln(n)\right) &\leq P\left(|C^{\geq a\ln(n)}| \geq 1\right) \\ &\leq E\left(|C^{\geq a\ln(n)}|\right) \\ &= nP(|C(v)| \geq a\ln(n)) \end{aligned} \tag{3.19}$$

[5]We have required a specific point in which the process of exploration starts, so that we are focusing only on the first component exploration. In this way it is possible to use the variable S_i and not S_i^1.

Finally, from equations (3.17), (3.18) and (3.19) we can conclude that

$$\begin{aligned} P(|C_{max}| \geq a\ln(n)) &\leq nP(|C(v)| \geq a\ln(n)) \\ &\leq ne^{-a\ln(n)I_P} \\ &= ne^{\ln(n)^{-aI_P}} \\ &= n^{1-aI_P} \\ &\underset{n\to\infty}{\to} 0 \end{aligned}$$

since $aI_P > 1$.

Proof of equation 3.4

The proof of this second equation is a little more complicated than the one just developed: for this reason, it is hard to give an outline of the procedure. However, after a few steps, it will be clear how the proof works.

Note that, from the proof of Theorem 3.3, it can be deduced that

$$P(|C_{max}| < b\ln(n)) = P\left(|C^{\geq b\ln(n)}| = 0\right), \tag{3.20}$$

and by Chebyshev inequality, it follows that

$$P\left(|C^{\geq b\ln(n)}| = 0\right) \leq \frac{Var\left(|C^{\geq b\ln(n)}|\right)}{E\left(|C^{\geq b\ln(n)}|\right)^2}. \tag{3.21}$$

At this point, there are two major tasks in this proof: find an upper bound for the variance and a lower bound for the expectation, both of them as a function of n, in order to find a new bound for $P(|C_{max}| < b\ln(n))$.

Firstly, we consider the bound of the variance, which is much more simpler than the one for the expectation. From equation (3.10)

$$Var\left(|C^{\geq b\ln(n)}|\right) \leq nE\left(|C(v)|\mathbf{1}_{\{|C(v)|\geq b\ln(n)\}}\right).$$

Now recall that the expectation of a discrete random variable X taking only non-negative values can also be defined as

$$E(x) = \sum_{i=1}^{\infty} P(X \geq i),$$

since

$$\begin{aligned} \sum_{i=1}^{\infty} P(X \geq i) &= \sum_{i=1}^{\infty}\sum_{j=i}^{\infty} P(X = j) = \sum_{j=1}^{\infty}\sum_{i=1}^{j} P(X = j) \\ &= \sum_{j=1}^{\infty} jP(X = j) = E(X). \end{aligned}$$

Thus, calling $k_n = [b\ln(n)]$ the integer part of $b\ln(n)$, we have that

$$\begin{aligned} Var\left(|C^{\geq b\ln(n)}|\right) &\leq n\sum_{i=k_n}^{n} P(|C(v)| \geq i) \\ &= n\sum_{i=k_n}^{n} P(|C(v)| > i-1) \\ &\leq n\sum_{i=k_n}^{n} e^{-(i-1)I_P}, \end{aligned}$$

where the last inequality follows from (3.17). Exploiting some of the properties of the sums, we can deduce that

$$\begin{aligned} \sum_{i=k_n}^{n} e^{-(i-1)I_P} &= \sum_{s=k_n-1}^{n-1} e^{-sI_P} = \sum_{s=0}^{n-1} e^{-sI_P} - \sum_{s=0}^{k_n-2} e^{-sI_P} \\ &= \frac{1-e^{-nI_P}}{1-e^{-I_P}} - \frac{1-e^{-(k_n-1)I_P}}{1-e^{-I_P}} \\ &= \frac{e^{-(k_n-1)I_P} - e^{-nI_P}}{1-e^{-I_P}} \\ &\leq \frac{e^{-(k_n-1)I_P}}{1-e^{-I_P}} \\ &\approx \frac{e^{-(b\ln(n)-1)I_P}}{1-e^{-I_P}} \\ &= \frac{n^{-bI_P}e^{-I_P}}{1-e^{-I_P}}. \end{aligned}$$

Thus, the upper bound for the variance is

$$Var\left(|C^{\geq b\ln(n)}|\right) \leq n^{1-bI_P}\frac{e^{I_P}}{1-e^{-I_P}}. \tag{3.22}$$

Now, we have to consider the expectation of $|C^{\geq b\ln(n)}|$. From equation (3.9) we know that

$$E\left(|C^{\geq b\ln(n)}|\right) = nP\left(|C(v)| \geq b\ln(n)\right).$$

This means that we can bound this expectation, by finding a lower bound for the probability that the size of a component of the random graph is bigger or equal to $b\ln(n)$. This is quite a hard task, which we will achieve according to the following steps:

1. The size of a component of the random graph will be compared to the progeny of a binomial branching process with parameters $(n-k_n, c/n)$, where $k_n = [b\ln(n)]$;
2. We will exploit the link between binomial and Poisson branching processes and the law of the mass distribution of the total progeny for a Poisson branching process, denoted with T_P, to compute $P(T_P \geq k_n)$;
3. Using Stirling's formula, we will be able to express the probability that $T_P \geq k_n$, as a function of n, deriving in this way the desired lower bound.

We now start with the first point. To do this, let $\tau_k = \min\{i : N_i \leq n-k\}$, where N_i is the number of neutral vertices after i explorations, and consider a process of exploration starting from vertex v. As long as $S_i > 0$, where S_i denotes the number of active vertices after the i-th exploration, we are exploring the component of v, which is the first one. If $S_i > 0$, for every $i \in [0, k-1]$, then $\tau_k \leq k-1$, since, after $k-1$ explorations, there are at least one active vertex and $k-1$ inactive vertices. At τ_k, there are are at least $k - \tau_k$ active vertices, since $S_{\tau_k} = n - \tau_k - N_{\tau_k} \geq k - \tau_k$. Moreover, $S_i > 0$ for at least a further $k - \tau_k - 1$ steps, since S_i decreases by at most a unity at every step. Hence

$$P(S_i > 0, 0 \leq i \leq k-1) = P(S_t > 0, 0 \leq i \leq \tau_k, \tau_k \leq k-1).$$

Now, let $X_1^{bp}, \ldots, X_i^{bp} \overset{i.i.d.}{\sim} B(n-k, c/n)$ and, for $i \leq \tau_k - 1$, let also $Y_1, \ldots, Y_i$ be a sequence of binomial random variables with parameters $(N_{i-1} - (n-k), c/n)$; then

$$X_i = X_i^{bp} + Y_i \tag{3.23}$$

is a binomial random variable with parameters $(N_{i-1}, c/n)$, which dominates X_i. Moreover, the sequence $X_1, \ldots, X_i$, as defined in (3.23) represents the exploration of the first component of a random graph, as long as $i \leq \tau_k - 1$ and $\tau_k \leq k-1$. Now note that, calling $S_i^{bp} = X_1^{bp} + \cdots + X_i^{bp}$,

$$\begin{aligned} P(|C(v)| \geq k) &= P(S_i > 0, 0 \leq i \leq \tau_k, \tau_k \leq k-1) \\ &\geq P(S_i^{bp} > 0, 0 \leq i \leq \tau_k, \tau_k \leq k-1) \\ &\geq P(S_i^{bp} > 0, 0 \leq i \leq k-1, \tau_k \leq k-1) \\ &= P(S_i^{bp} > 0, 0 \leq i \leq k-1) = P(T^B_{n-k,c/n} \geq k), \end{aligned}$$

since $\{S_i^{bp} > 0, 0 \leq i \leq k-1\} \Rightarrow \{S_i > 0, 0 \leq i \leq k-1\} \Rightarrow \{\tau_k \leq k-1\}$, and where $T^B_{n-k,c/n}$ denotes the total progeny of a binomial branching process with parameters $(n-k, c/n)$. Since the exploration of the random graph can start from any vertex, the result is true for any vertex v. Thus, we can write

$$P\left(|C(v)| \geq k_n\right) \geq P(T^B_{n-k_n,c/n} \geq k_n). \tag{3.24}$$

At this point (the second one of the list above), we can exploit the relationship between Poisson and binomial branching processes. If we let $T^P_{c_n}$ be total progeny of a Poisson branching process with mean $c_n = c\frac{n-k_n}{n}$, then, by Theorem 2.10, it follows that

$$P(T^B_{n-k_n} \geq k_n) = P(T^P_{c_n} \geq k_n) + o(1).$$

Moreover, because of Theorem 2.9, we have that

$$P(T^P_{c_n} \geq k_n) = \sum_{s=k_n}^{\infty} P(T^P_{c_n} = s) = \sum_{s=k_n}^{\infty} \frac{e^{-sc_n}(sc_n)^{s-1}}{s!}.$$

We now introduce Stirling's formula. In its most common formulation, this equation states that

$$\lim_{n\to\infty} \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{n!} = 1,$$

which can also be written as,

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n}(1+o(1)).$$

Now, using Stirling's formula and recalling all the results obtained in this part of the proof, we can notice that, calling I_{P_n} the rate function of a Poisson random variable with mean c_n,

$$\begin{aligned} P(|C(v)| \geq k_n) &\geq \sum_{s=k_n}^{\infty} \frac{e^{-sc_n}(c_n)^{s-1}e^s}{\sqrt{2\pi s^3}}(1+o(1)) \\ &= \frac{1}{c_n}\sum_{s=k_n}^{\infty} \frac{e^{-sI_{P_n}}}{\sqrt{2\pi s^3}}(1+o(1)) \\ &\geq \frac{1}{c_n}\sum_{s=k_n}^{\infty} \frac{e^{-sI_{P_n}}}{\sqrt{2\pi k_n^3}}(1+o(1)) \end{aligned}$$

since

$$e^{-sI_{P_n}} = e^{-s(c_n-1-\ln(c_n))} = e^{-sc_n}e^{-s}c_n^s.$$

Moreover, we can notice that $c_n < 1$, since $c < 1$ and $\frac{n-k_n}{n} < 1$, and that $k_n = b\ln(n) > 1$ for n large enough, so that

$$\begin{aligned} P(|C(v)| \geq k_n) &\geq \frac{1}{c_n}\sum_{s=k_n}^{\infty} \frac{e^{-sI_{P_n}}}{\sqrt{2\pi k_n^3}}(1+o(1)) \\ &\geq e^{-k_n I_{P_n}}(1+o(1)). \end{aligned}$$

Let's consider the rate function I_{P_n}: we analyze now its relationship with I_c. By definition,

$$\begin{aligned} I_{P_n} &= c_n - 1 - \ln(c_n) \\ &= c\frac{n-k_n}{n} - 1 - \ln\left(c\frac{n-k_n}{n}\right) \\ &= c\frac{n-k_n}{n} - 1 - \ln(c) + \ln\left(\frac{n-k_n}{n}\right) \\ &\underset{n\to\infty}{\to} c - 1 - \ln(c) = I_P. \end{aligned}$$

This result can also be written as $I_{P_n} = I_P + o(1)$. Thus,

$$\begin{aligned} P(|C(v)| \geq k_n) &\geq e^{-k_n I_{P_n}}(1+o(1)) \\ &\approx e^{-b\ln(n)I_P(1+o(1))} \\ &= n^{-bI_P(1+o(1))}. \end{aligned}$$

In this way, we have found the lower bound for the expectation of $|C^{\geq b\ln(n)}|$, which is

$$E\left(|C^{\geq b\ln(n)}|\right) = nP\left(|C(v)| \geq b\ln(n)\right) \geq n^{(1-bI_P)(1+o(1))} \geq n^{\alpha}, \qquad (3.25)$$

for any $\alpha \in (0, 1-aI_P)$, since $bI_P < 1$. Finally, recalling equations (3.20), (3.21) and the two bounds obtained in (3.22) and (3.25), we have that

$$P(|C_{max}| < b\ln(n)) \leq n^{1-bI_P-2\alpha}\frac{e^{I_p}}{1-e^{I_P}} \underset{n\to\infty}{\to} 0,$$

when we choose an α such that $2\alpha > 1 - bI_P$.

3.5 The supercritical regime

We now prove equation (3.5) from Theorem 3.1, which is also known as the *law of large numbers for the giant component*. The proof consists of three main parts, which are now introduced.

- The first one shows that there are no components with size between $k_n = [a \ln(n)]$ and αn, with $\alpha < \beta_c$, that is

$$\lim_{n\to\infty} P(\exists v : k_n \leq |C(v)| \leq \alpha n) = 0; \tag{3.26}$$

- We will prove a slight variation of the main equation, which uses $|C^{\geq k_n}|$ instead of the size of the largest component. The result will be

$$\lim_{n\to\infty} P(|\ |C^{\geq k_n}| - \beta_c n| \leq n^v) = 1 \tag{3.27}$$

- The final step will be to use the two previous results, (3.26) and (3.27), to deduce the law of large numbers for the giant component;

Actually, we are going to start from the final step, assuming that (3.26) and (3.27) hold.

Note that it is always true that

$$|C_{max}| \leq |C^{\geq k_n}| \tag{3.28}$$

as long as there exists a component of size bigger than k_n: but this is certainly implied by (3.27). However, if $|C_{max}| < |C^{\geq k_n}|$, there are at least two components with at least k_n vertices. Moreover, they must have more than αn vertices, since (3.26) is assumed to be true. Thus

$$|C^{\geq k_n}| > 2\alpha n. \tag{3.29}$$

We also know that $|C^{\geq k_n}| \leq \beta_c n + n^v$, since, from (3.27)

$$\begin{aligned} P(|\ |C^{\geq k_n}| - \beta_c n| \leq n^v) &= P(-n^v \leq |C^{\geq k_n}| - \beta_c n \leq n^v) \\ &\leq P(|C^{\geq k_n}| \leq \beta_c n + n^v) \end{aligned}$$

for large values of n. However, when $2\alpha > \beta_c$, (3.29) is in contradiction with

$$|C^{\geq k_n}| \leq \beta_c n + n^v < 2\alpha n + o(n).$$

Thus, since (3.28) holds, we must have that

$$|C_{max}| = |C^{\geq k_n}|$$

conditional on (3.26) and (3.27), from which the result follows.

So, we now have to prove the two conditions from which the result is deduced and we start with (3.26). Note that, by Markov inequality

$$\begin{aligned}
P(\exists v : k_n \leq |C(v)| \leq \alpha n) &= P(|C^{\geq k_n}| - |C^{\geq \alpha n+1}| \geq k_n) \\
&\leq P(|C^{\geq k_n}| - |C^{\geq \alpha n+1}| \geq 1) \\
&\leq E(|C^{\geq k_n}| - |C^{\geq \alpha n+1}|) \\
&= E(|C^{\geq k_n}|) - E(|C^{\geq \alpha n+1}|) \\
&= n\left(P(|C(v)| \geq k_n) - P(|C(v)| \geq \alpha n + 1)\right) \\
&= nP(k_n \leq |C(v)| \leq \alpha n) \\
&= n \sum_{i=k_n}^{\alpha n} P(|C(v)| = i) \\
&= n \sum_{i=k_n}^{\alpha n} P(\inf\{i : S_i = 0\}) \\
&\leq n \sum_{i=k_n}^{\alpha n} P(S_i = 0),
\end{aligned}$$

where S_i denotes the number of active vertices after the i-th exploration of a process starting at vertex v. Thus, in order to proceed we have to find the distribution of the number of active vertices. We can do that, by finding the distribution of N_i, the number of neutral vertices after the i-th exploration, as will become apparent later. Recall that, by the random walk construction,

$$N_i = n - i - S_i = n - i - S_{i-1} - X_i + 1,$$

where $X_i \sim B(N_{i-1}, c/n)$. Thus

$$N_i = N_{i-1} - X_i \sim Y_i,$$

where $Y_i \sim B(N_{i-1}, 1 - c/n)$. This result is true since, if X is a $\mathrm{B}(m, p)$ and $Y = m - X$, with $m \geq 1$ and $p \in [0, 1]$, then Y is a $\mathrm{B}(m, 1 - p)$. To understand this fact, note that the probability of a certain number of successes $P(Y = y)$ for Y is equivalent to the probability of the same number of failures for X, since

$$P(Y = y) = P(m - X = y) = P(X = m - y).$$

Therefore, in order to obtain the same random variable is sufficient to switch the value of the parameter. Thus, we have that $N_i \sim Y_i \sim B(N_{i-1}, 1 - c/n)$ and we will prove that $N_i \sim B(n - 1, (1 - c/n)^i)$. In order to do that, recall that if $N \sim B(n, p)$ and $S \sim B(N, q)$, then, conditionally on N, $S \sim B(n, pq)$. In order to understand this statement we can think of N like n coin tosses with probability of success p and, for each trial, we throw an other coin with success probability q. The amount of successes of S will be the cases in which both trials ended in success and this corresponds to throwing a coin with success probability pq n times. Now, considering N_i, we have that $N_0 = n - 1$, thus $N_1 \sim B(n - 1, 1 - c/n)$ and $N_2 \sim B(N_1, 1 - c/n) \sim B(n - 1, (1 - c/n)^2)$: by induction, the result easily follows.

Now, since $S_i = n - i - N_i = n - 1 - (i - 1) - N_i$, then

$$S_i + (i - 1) = n - 1 - N_i \sim B(n - 1, 1 - (1 - c/n)^i)$$

In this way we have obtained the distribution of S_i. Now let $R \sim B(n-1, 1-(1-c/n)^i)$, $C \sim B(n, 1-(1-c/n)^i)$ and $X \sim B(n, 1-e^{-ic/n})$, then

$$\begin{aligned} P(S_i = 0) &= P(R = i-1) \leq P(R \leq i-1) \\ &\leq P(C \leq i) \leq P(X \leq i), \end{aligned}$$

where the last inequality follows from the Taylor approximation in (1.8). Then, since the exponential is an invertible transformation and applying Markov inequality, we have that, for $s \geq 0$,

$$\begin{aligned} P(S_i = 0) &\leq P(-X \geq -i) = P(e^{-sX} \geq e^{-si}) \\ &\leq e^{si} E(e^{-sX}) = e^{si} \sum_{j=0}^{n} e^{-sj} P(X = j) \\ &= e^{si} \sum_{j=0}^{n} e^{-sj} \binom{n}{j} \left(1 - e^{-\frac{ic}{n}}\right)^j \left(e^{-\frac{ic}{n}}\right)^{n-j} \\ &= e^{si} \sum_{j=0}^{n} \binom{n}{j} \left(e^{-s}\left(1 - e^{-\frac{ic}{n}}\right)\right)^j \left(e^{-\frac{ic}{n}}\right)^{n-j} \\ &= e^{si} \left(e^{-s}\left(1 - e^{-\frac{ic}{n}}\right) + e^{-\frac{ic}{n}}\right)^n \\ &= e^{si} \left(1 - \left(1 - e^{-\frac{ic}{n}}\right)\left(1 - e^{-s}\right)\right)^n \\ &\leq e^{si - n\left(1 - e^{-ic/n}\right)\left(1 - e^{-s}\right)} \end{aligned} \tag{3.30}$$

The last inequality is also true for the minimum of the exponent of the right-hand side of (3.30): we now compute its derivative in s and we set it equal to zero.

$$\frac{d}{ds}\left[si - n\left(1 - e^{-ic/n}\right)\left(1 - e^{-s}\right)\right] = i - n\left(1 - e^{-ic/n}\right)e^{-s} = 0.$$

Thus,

$$\begin{aligned} e^{-s^*} &= \frac{i}{n\left(1 - e^{-ic/n}\right)} \\ \Longrightarrow s^* &= \ln\left(\frac{n\left(1 - e^{-ic/n}\right)}{i}\right), \end{aligned}$$

where s^* is the minimized value (it is straightforward to check the second derivative). We now substitute s^* in the exponent of (3.30), so that, calling $g(i) = \frac{n\left(1-e^{-ic/n}\right)}{i}$,

$$\begin{aligned} s^* i - n\left(1 - e^{-ic/n}\right)\left(1 - e^{-s^*}\right) &= \\ &= i \ln(g(i)) - n\left(1 - e^{-ic/n}\right)\left(1 - \frac{i}{n\left(1 - e^{-ic/n}\right)}\right) = \\ &= i \ln(g(i)) - n\left(1 - e^{-ic/n}\right) + i = \\ &= -i\left(g(i) - 1 - \ln(g(i))\right) = -i I_{P(g(i))}(1), \end{aligned}$$

where $I_{P(g(i))}(1)$ is the rate function in 1 of a Poisson random variable with mean $g(i) > 0$. We need $I_{P(g(i))}(1)$ to be bigger than zero and this happens if $g(i) > 1$. Thus, we now have to show that $g(i) > 1$ for some values of i.

If $\frac{i}{n} < \beta_c$, then $g(i) > 1$, since

$$\begin{aligned} g(i) &= \frac{n\left(1 - e^{-\frac{ic}{n}}\right)}{i} \\ &\geq \frac{n\left(1 - e^{-c\beta_c}\right)}{i} \\ &\geq \frac{nc\beta_c}{i} \geq \frac{n\beta_c}{i} > 1. \end{aligned}$$

Moreover, since $\alpha < \beta_c$, then $g(i) > 1$ for all $k_n \leq i \leq \alpha n$.

Thus, $I_{P(g(i))}(1) > 0$ for $k_n \leq i \leq \alpha n$: so we can bound it below by a strictly positive constant $K(\alpha, c)$. Now, summarizing the main steps developed in this part of the proof, we have that

$$\begin{aligned} P(\exists v : k_n \leq |C(v)| \leq \alpha n) &\leq n \sum_{i=k_n}^{\alpha n} P(S_i = 0) \\ &\leq n \sum_{i=k_n}^{\alpha n} e^{s*i - n\left(1-e^{-ic/n}\right)\left(1-e^{-s*}\right)} \\ &= n \sum_{i=k_n}^{\alpha n} e^{-iI_{P(g(i))(1)}} \\ &\leq n \sum_{i=k_n}^{\alpha n} e^{-iK(\alpha,c)} \\ &= n \frac{e^{-k_n K(\alpha,c)}}{1 - e^{-K(\alpha,c)}} \\ &\approx n \frac{e^{-a\ln(n)K(\alpha,c)}}{1 - e^{-K(\alpha,c)}} \overset{n\to\infty}{\to} 0, \end{aligned}$$

for values of a large enough.

We now consider the other condition from which the main result follows, that is (3.27). In order to prove it, we first have to relate β_c to the expectation of $|C^{\geq k_n}|$. Recall that (3.9) states that

$$E(|C^{\geq k_n}|) = nP(|C(v)| \geq k_n).$$

The probability that the size of a component is bigger than k_n in the supercritical phase is comparable to the survival probability of a Poisson branching process. More formally we are going to show that

$$P(|C(v)| \geq k_n) = \beta_c + O(k_n/n) \tag{3.31}$$

In order to prove (3.31), we have to find an upper and a lower bound for $P(|C(v)| \geq k_n)$ that coincides with $\beta_c + O(k_n/n)$. We can find the upper bound

using the total progeny of a binomial branching process, as done in (3.16), and then comparing it to a Poisson branching process by Theorem 2.10, to obtain

$$P(|C(v)| \geq k_n) \leq P(T^B_{n,c/n} \geq k_n) = P(T^P_c \geq k_n) + O(k_n/n).$$

Then, it follows from Theorem 2.8 that

$$\begin{aligned} P(T^P_c \geq k_n) &= P(T^P_c = \infty) + P(k_n \leq T^P_c < \infty) \\ &= \beta_c + O(e^{-k_n I_P}) \\ &\approx \beta_c + O(e^{-a\ln(n) I_P}) \\ &= \beta_c + O(n^{-aI_P}) \\ &= \beta_c + o(1), \end{aligned}$$

for $a > 0$, since $I_P > 0$. Thus

$$P(|C(v)| \geq k_n) \leq \beta_c + O(k_n/n). \tag{3.32}$$

It is possible to find as well a lower bound for $P(|C(v)| \geq k_n)$ using, in this case, the result obtained in (3.24) and then, again, Theorem 2.10, to obtain

$$P(|C(v)| \geq k_n) \geq P(T^B_{n-k_n,c/n} \geq k_n) = P(T^P_{c_n} \geq k_n) + O(k_n/n).$$

Again, we use Theorem 2.8, to compute $P(T^P_{c_n} \geq k_n)$.

$$\begin{aligned} P(T^P_{c_n} \geq k_n) &= P(T^P_{c_n} = \infty) + P(k_n \leq T^P_{c_n} < \infty) \\ &= \beta_{c_n} + O(e^{-k_n I_{P_n}}) \\ &= \beta_{c_n} + o(1), \end{aligned}$$

where β_{c_n} is the survival probability of a Poisson Branching Process with mean $c_n = c\frac{n-k_n}{n}$. We now have to relate β_{c_n} to β_c in order to obtain the same upper and lower bound. However, since $k_n = o(n)$, by definition of c_n, we have that

$$\beta_{c_n} = \beta_c + o(1)$$

Thus, we can conclude that

$$P(|C(v)| \geq k_n) \geq \beta_c + O(k_n/n),$$

which proves, together with (3.32), equation (3.31). In this way, we deduce that

$$E(|C^{\geq k_n}|) = nP(|C(v)| \geq k_n) = n\beta_c + nO(k_n/n) = n\beta_c + O(k_n).$$

We can now consider again equation (3.27) that we are trying to prove. Indeed

$$\begin{aligned} P(|\,|C^{\geq k_n}| - \beta_c n| \leq n^v) &\geq P(|\,|C^{\geq k_n}| - \beta_c n| \leq n^v/2) \\ &= P(|\,|C^{\geq k_n}| - E(|C^{\geq k_n}|) + O(k_n)| \leq n^v/2). \end{aligned}$$

Since $k_n = o(n^v)$ for any $v \in (1/2, 1)$, we can simply consider $P(|\,|C^{\geq k_n}| - E(|C^{\geq k_n}|)| \leq n^v/2)$, which can be bounded using (3.11) and Chebyshev inequality. Thus,

$$\begin{aligned} P(|\,|C^{\geq k_n}| - E(|C^{\geq k_n}|)| \leq n^v/2) &= 1 - P(|\,|C^{\geq k_n}| - E(|C^{\geq k_n}|)| > n^v/2) \\ &\geq 1 - \frac{Var(|C^{\geq k_n}|)}{4n^{2v}} \\ &\geq 1 - \frac{(ck_n+1)nE(|C(v)|\mathbf{1}_{\{|C(v)|<k_n\}})}{4n^{2v}} \\ &= 1 - \frac{(ck_n+1)E(|C(v)|\mathbf{1}_{\{|C(v)|<k_n\}})}{4n^{2v-1}}. \end{aligned}$$

Now note that, $E(|C(v)|\mathbf{1}_{\{|C(v)|<k_n\}})$ can be bounded above by k_n, because of the presence of the indicator function that forces the component to have size less than k_n. Indeed,

$$\begin{aligned} P(|\ |C^{\geq k_n}| - \beta_c n| \leq n^v/2) &\geq 1 - 4n^{1-2v}(ck_n^2 + k_n) \\ &\underset{n\to\infty}{\to} 1, \end{aligned}$$

since $v > 1/2$ and $k_n = o(n^\delta)$, with $\delta = 1 - 2v$.

3.6 The critical case

We now study what happens in the case $c = 1$, which is also known as the *critical value*. Let's start with a simulation of a random graph with 1000 vertices and edge probability 1/1000. Figure 3.4 shows a graph in which there are some components of medium dimension, bigger than the ones in the subcritical regime, but much more smaller than the giant component of the supercritical phase. In particular, we can see a component with red vertices, the largest one, in the middle of the graph, which is of discrete dimension and bigger than all the others. However, the difference between the largest and the other components is not so pronounced as in the supercritical phase. Moreover, even the components which are not the largest one, have a bigger size than the largest component in the subcritical phase.

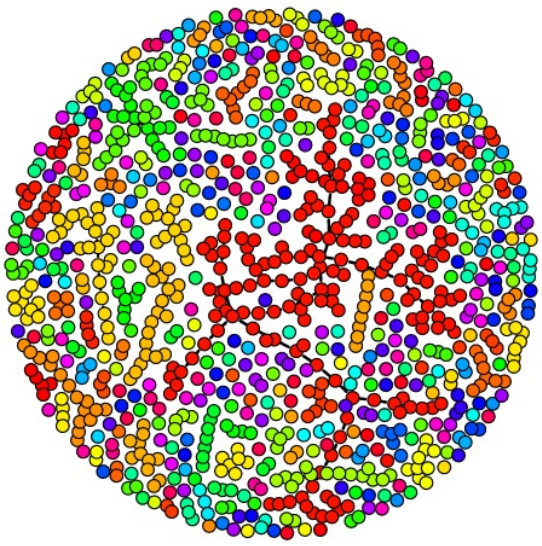


Figure 3.4: Critical Random Graph

Table 3.3 shows the size of the four largest components in the simulated random graph and confirms the observations derived from Figure 3.4.

In particular, the main interesting feature of the table is the number of vertices in the largest component is comparable to the value $n^{2/3} = 100$, which,

1st	2nd	3rd	4th
143	23	20	16

Table 3.3: Size of the Components in the Critical Case.

indeed is the order of the size of C_{max} in the critical phase. This fact is introduced in the following theorem.

Theorem 3.4. *Let $G(n, c/n)$ be a random graph with $c = 1 + rn^{-1/3}$ and $r \in \mathbb{R}$. Let also $b > 0$, then, for w sufficiently large,*

$$P\left(\frac{n^{2/3}}{w} \leq |C_{max}| \leq wn^{2/3}\right) \geq 1 - \frac{b}{w}. \tag{3.33}$$

We can notice that the mean number of neighbors of a vertex in Theorem 3.33 is not precisely 1: however, for graphs with a high number of vertices, it tends to one. The meaning of the previous theorem is that, with a probability close to 1, the size of the largest component is in an interval which includes the value $n^{2/3}$. The "proximity" of the probability to one depends on the width of the interval. Note that the conclusion of this theorem is a little bit different than the ones of Theorem 3.1: indeed, in Theorem 3.1 we proved that the probability of the size of the largest components not to be of a certain order tends to zero. In this case, on the contrary, we are not proving any asymptotic result: Theorem 3.33 "just" states that the probability of the size to be of the order $n^{2/3}$ is "substantially" 1.

We now start developing a proof of Theorem 3.33: however, at some stages, we will not prove every single passage, since we want to give just a justification of this result. The proof consists of these main steps:

1. Find an upper and a lower bound for $P(|C(v)| \geq k)$ using the relationship with branching processes;
2. Exploiting the relationship between C_{max} and $C^{\geq k}$ and the upper bound found in the first point, to derive an upper bound for $P(|C(v)| > wn^{2/3})$;
3. Using the lower bound of the first point of this summary, we will find an upper bound for $P(|C(v)| < w^{-1}n^{2/3})$;
4. Using the results of point 2 and 3, we will be able to derive the statement of Theorem 3.33.

The first point will be the one in which we will not develop every step of the proof. It is based on the relationship that we have used several times during this dissertation with branching processes. For the upper bound, recall that, by equation (3.16) and by Theorem 2.10, we have that

$$P(|C(v)| \geq k) \leq P(T^B_{n,c/n} \geq k) = P(T^P_c \geq k) + o(1).$$

On the contrary, for the lower bound, using equation (3.24) and again Theorem 2.10, we have that

$$P(|C(v)| \geq k) \geq P(T^B_{n-k,c/n}) = P(T^P_{c_1} \geq k) + o(1),$$

where $c_1 = c(n-k)/n$. At this point, it is possible, exploiting some properties of the branching processes that were not introduced in Chapter 2, to find specific bounds for the critical phase of $P(T_c^P \geq k) + o(1)$ and $P(T_{c_1}^P \geq k) + o(1)$. However, we will not discuss the derivation of these bounds here. Indeed, it is possible to prove that

$$\frac{d_1}{\sqrt{k}} \leq P(|C(v)| \geq k) \leq d_2 \left(|r| n^{-1/3} + \frac{1}{\sqrt{k}} \right). \tag{3.34}$$

We can now consider the second point of the summary. By the relationship between C_{max} and $C^{\geq k}$ and by Markov inequality, we have that

$$P(|C_{max}| > wn^{2/3}) = P(|C^{\geq wn^{2/3}}| \geq wn^{2/3}) \leq \frac{E(|C^{\geq wn^{2/3}}|)}{wn^{2/3}}.$$

Now, by equations 3.34 and 3.9, we can deduce that

$$\begin{aligned} E(|C^{\geq wn^{2/3}}|) = nP(|C(v)| \geq wn^{2/3}) &\leq nd_2 \left(|r| n^{-1/3} + \frac{1}{\sqrt{wn^{2/3}}} \right) \\ &= nd_2 n^{-1/3} \left(|r| + \frac{1}{\sqrt{w}} \right) \\ &= n^{2/3} d_2 \left(|r| + \frac{1}{\sqrt{w}} \right). \end{aligned}$$

Thus,

$$\begin{aligned} P(|C_{max}| \geq wn^{2/3}) &\leq \frac{n^{2/3} d_2 \left(|r| + \frac{1}{\sqrt{w}} \right)}{wn^{2/3}} \\ &= \frac{d_2}{w} \left(|r| + \frac{1}{\sqrt{w}} \right) \\ &\leq 2 \frac{d_2 |r|}{w}. \end{aligned} \tag{3.35}$$

Fort the third point, if we let $\hat{r} = \max\{|r|, 1\}$ and C_{max}^1 be the largest component size of a random graph with edge probability $\frac{c_1}{n} = \frac{1 - \hat{r} n^{-1/3}}{n}$, then we have that

$$P(|C_{max}| < w^{-1} n^{2/3}) \leq P(C_{max}^1 < w^{-1} n^{2/3}),$$

since $c_1 < c$. Moreover, since $-\hat{r} \leq -1$, we can restrict the proof to values of $r \in (-\infty, 1]$. Recalling the relationship with $|C^{\geq k}|$ and Chebyshev inequality we can note that

$$P(|C_{max}| < w^{-1} n^{2/3}) = P(|C^{\geq w^{-1} n^{2/3}}| = 0) \leq \frac{Var\left(|C^{\geq w^{-1} n^{2/3}}| \right)}{E\left(|C^{\geq w^{-1} n^{2/3}}| \right)^2}.$$

Thus, we have to find an upper bound for the variance and a lower bound for the expectation. We start with the expectation and, by equation (3.34), we deduce that

$$E\left(|C^{\geq w^{-1} n^{2/3}}| \right) = nP(|C(v)| \geq w^{-1} n^{2/3}) \geq n \frac{d_1}{\sqrt{w^{-1} n^{2/3}}} = d_1 \sqrt{w} n^{2/3}. \tag{3.36}$$

Now, we focus on the bound for the variance, and recalling the one obtained in (3.10), it follows that

$$Var\left(|C^{\geq w^{-1}n^{2/3}}|\right) \leq nE(|C(v)|\mathbf{1}_{\{|C(v)|\geq w^{-1}n^{2/3}\}}) \leq nE(|C(v)|).$$

Recalling that $P(|C(v)| > k) \leq P(T^B_{n,c/n})$, one can easily deduce that it also holds $E(|C(v)|) \leq E(T)$. Thus, from equation 2.3, we have that, since $r < -1$,

$$E(T) = \frac{1}{1-(1-rn^{-1/3})} = \frac{n^{1/3}}{|r|},$$

from which we can deduce that

$$Var\left(|C^{\geq w^{-1}n^{2/3}}|\right) \leq \frac{n^{4/3}}{|r|} \leq n^{4/3}. \tag{3.37}$$

In this way we have found the upper bound for $P(|C_{max}| > w^{-1}n^{2/3})$, which is

$$P(|C_{max}| > w^{-1}n^{2/3}) \leq \frac{n^{4/3}}{(d_1\sqrt{w}n^{2/3})^2} = \frac{1}{d_1^2 w}.$$

Finally, for the last point of the summary, we can conclude that

$$\begin{aligned} P\left(\frac{n^{2/3}}{w} \leq |C_{max}| \leq wn^{2/3}\right) &= 1 - P\left(|C_{max}| > wn^{2/3}\right) \\ &\quad -P\left(|C_{max}| < w^{-1}n^{2/3}\right) \\ &\geq 1 - \frac{1}{d_1^2 w} - \frac{2d_2|r|}{w} = 1 - \frac{b}{w}, \end{aligned}$$

with $b = d_1^{-2} + 2d_2|r|$.

Bibliography

[1] Robin Wilson and John Watkins. *Graphs: An Introductory Approach.* Wiley, New York, 1990.

[2] Reinard Diestel. *Graph Theory.* Springer-Verlag, New York, 2000.

[3] Svante Janson, Tomasz Luczak, and Andrzej Rucinski. *Random Graphs.* Wiley, New York, 2000.

[4] Frank den Hollander. *Large Deviations.* AMS, Providence, 2000.

[5] Remco van der Hofstad. Random graphs and complex networks. Lecture Notes, 2011.

[6] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times.* AMS, Providence, 2009.

[7] Hermann Thorisson. *Coupling, Stationarity and Regeneration.* Springer-Verlag, New York, 2000.

[8] Charles M. Grinstead and J. Laurie Snell. *Introduction to Probability.* AMS, Providence, 1998.

[9] Remco van der Hofstad and M. Keane. An elementary proof of the hitting time theorem. *The American Mathematical Monthly*, 2008.

[10] Geoffrey Grimmett and Dominc Welsh. *Probability: An Introduction.* Clarendon Press, Oxford, 1986.

[11] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms.* McGraw Hill, Cambridge, 2001.

[12] Svante Janson and Malwina J. Luczak. A new approach to the giant component problem. *Random Structures and Algorithms*, 2009.