BACHELOR'S THESIS

Degree in Mathematics

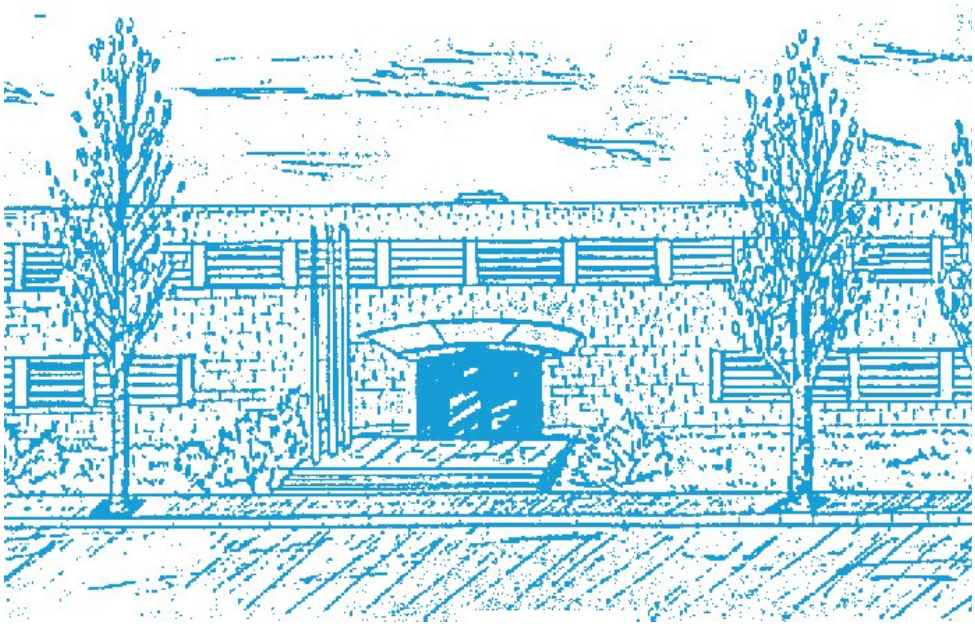
Title: Threshold phenomena in random graphs

Author: Nil Redón Orriols

Advisor: Juan José Rué Perna

Department: Mathematics

Academic year: 2018-2019



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

Universitat Politècnica de Catalunya
Facultat de Matemàtiques i Estadística

Degree in Mathematics

Bachelor's Degree Thesis

Threshold phenomena in random graphs

Nil Redón Orriols

Supervised by Juan José Rué

June 2019

Abstract

In the 1950s, random graphs appeared for the first time in a result of the prolific hungarian mathematician Pál Erdős. Since then, interest in random graph theory has only grown up until now. In its first stages, the basis of its theory were set, while they were mainly used in probability and combinatorics theory. However, with the new century and the boom of technologies like the World Wide Web, random graphs are even more important since they are extremely useful to handle problems in fields like network and communication theory. Because of this fact, nowadays random graphs are widely studied by the mathematical community around the world and new promising results have been recently achieved, showing an exciting future for this field. In this bachelor thesis, we focus our study on the threshold phenomena for graph properties within random graphs.

Keywords and AMS Subject Classification

Random graphs, threshold phenomena, giant component.
Main AMS subjects: 05C80, 60B99, 05D40

Contents

1 Introduction **3**

2 Background and definitions **6**
2.1 Asymptotic and set theory notation 6
2.2 Graph Theory 6
2.3 Probability Theory 7

3 Random graph models **9**
3.1 The origins of random graphs 9
3.2 Definition of $G(n,p)$ and $G(n,M)$ 13
3.3 Graph properties. Asymptotic equivalence 14

4 Properties of $G(n,p)$: Threshold functions and subgraphs **17**

5 The birth of the giant component **28**
5.1 *DFS* approach 28
5.2 Galton-Watson approach 35

6 Hamiltonicity of $G(n,p)$ **42**
6.1 Main result 43
6.2 Proof of Theorem 6.2. Use of expanders. 46

7 Conclusions and future work **52**

8 Acknowledgements **54**

1. Introduction

Random graph models are today one of the most important research field in discrete mathematics, combinatorics and theoretical computer science. They were first introduced by Erdős and Rényi in 1959, using them in a proof of graph theory. Hence, they realized random graph were so interesting to study them as a separate area, and they published the groundbreaking paper *On random graphs* [ER59]. It launched the random graph theory, since it was the first work dedicated to this topic, so it has been extremely important to encourage the research and to develop the knowledge on this field. Random graphs are in the intersection of probability and graph theory, and nowadays they are basic to model very common situations in real life, such as social or computer networks or biological phenomena.

We focus in the Erdős-Rényi model for random graph generation. It assumes independence for all nodes when joining any other node with an edge. We can split this model into two different views, one came from Erdős and Rényi and the other from Gilbert. From a probabilistic point of view, Gilbert introduced $G(n, p)$, defined as a non directed graph with n vertices, such that for every two vertices exists an edge joining them with probability p. Looking from the combinatorics side, Erdős and Rényi defined $G(n, M)$ as a graph chosen in an uniformly random way in the set of graphs with n vertices and M edges. Both model different processes but they are similar structures. Indeed, we prove that the two models are equivalent under certain assumptions.

Another fundamental concept in all this work are threshold functions. A threshold function (or simply a threshold) for a mathematical model is informally defined as a function (related with the parameters of the model) such that a certain property is accomplished with probability tending to 1 if we are considering functions above the threshold, and is not (it has probability tending to 0) when we consider functions below the threshold. We have to think in the notions of above and below asymptotically. More precisely (the formal definitions are introduced in the notation chapter), if f is a threshold for a certain property, we say g is below f if and only if $g = o(f)$, and we say h is above f if and only if $h = O(f)$.

In this work, we are going to study random graph models focusing in the threshold phenomena for their properties. Many questions or events that appear from fields like probability theory, statistics, complexity theory or physics can be modeled as properties for random graphs. Given a property, if we are able to find that there is a threshold, we can find it explicitly and we can measure how fast is the change, then we reach a very good understanding of it. In the case of $G(n, p)$, taking $p = f(n)$ we prove that the most studied properties (indeed monotone properties, such as connectivity) have a threshold, and we study how a threshold could behave. The special structure of random graphs allows us to perform this study, as it is extremely difficult to achieve similar results for any general graph.

We work more in detail some particular cases in order to better understand this fact. We take a closer look at the threshold function for the property of having a concrete subgraph, for different cases of subgraphs. For example, we prove the threshold function when the subgraph we are looking for is the complete graph K_k. In a random graph $G(n,p)$ we ensure that there exists a subgraph which is K_k with high probability when $p = p(n)$ is above a given threshold. However, this problem is extremely difficult for a graph G. Also we work for more general subgraphs, focusing on balanced graphs. We also work in detail two particular cases of threshold functions for two well-known properties (hamiltonicity and connectivity) that require the introduction and usage of some interesting mathematical techniques.

The first particular case is called the birth of the giant component. It is a classic result in network theory that looks for the threshold for connected components of a graph. The concept of connected components for a graph rises immediately when we think that a graph is a representation of a real network. One of the first questions we ask ourselves given a network can easily be the following: Can I go from vertex v to vertex u in the graph? This simple example is equivalent to study the connected components for the graph, and its structure is one of the best indicators when thinking about how a graph looks like. We show two different and pretty recent proofs for this result, based on completely different concepts. The first one uses *Depth First Search* (*DFS*), an algorithm that can be used to search for connected components in a graph. The second approach relies on *Galton-Watson* processes, a stochastic branching process that leads to another different proof for the same main result. Moreover, every approach gives different and complementary results to the main one.

The second particular case we study is the threshold function for hamiltonicity. Recall that a graph is hamiltonian if it contains a hamiltonian cycle, this is a cycle which visits all vertices of the graph exactly once. Hamiltonicity has been widely studied among with several graph properties such as density, subgraphs and distance. Although many results are related with hamiltonicity of a graph (for example Dirac's and Ore's theorems, which roughly speaking state that a graph is Hamiltonian if it has enough edges), it is an extremely difficult problem to look for hamiltonicity on an arbitrary graph. As happened exactly when we introduced the case of finding a K_k as a subgraph of $G(n,p)$, a extremely difficult problem for a general graph G becomes much easier for $G(n,p)$. We claim, using techniques such as expanders and boosters, that for $p = p(n)$ over a threshold, $G(n,p)$ is *whp* hamiltonian.

As a summary, we set the following basic objectives for our work:

- Understanding random graph models, how they were originated and its utility.
- Understanding what is a threshold in random graph, and studying different cases of threshold.
- Working the particular example of the birth of the giant component, showing two different approaches to prove the result.

- Working the threshold for hamiltonicity, understanding the new concepts introduced

During all the research we try to make clear why we do everything, meaning that we try to relate all issues we see with applications in other mathematical fields or even with other different areas of knowledge. We try to follow a logic structure, meaning that the results and concepts we show are presented in a coherent order of difficulty and importance. We organize our work using the following structure, by chapter:

1. We introduce some concepts and notations of probability and graph theory that we use during all the work.
2. We look at the first appearance of the random graphs, trying to explain why they are so useful in many fields.
3. We present the basic random graph models $G(n, p)$ and $G(n, M)$, and the equivalences between both of them.
4. We study the properties of the most basic random graph model, namely $G(n, p)$.
5. We look at the particular case of the threshold function for the birth of the giant component in a graph, showing two completely different approaches to reach the same result, but with some interesting details in each case. [KS13]
6. We look at the threshold function for the hamiltonicity in a random graph. This is a recent result that also involves powerful tools such as expanders. [Kri16]
7. Finally, we extract some conclusions about each chapter and possible future work for further research.

2. Background and definitions

In this chapter, let us introduce to the reader, or recall if they are already known, some concepts, results and notation (from different fields like asymptotic, probability, graph and set theory) which are used in the work. If any concept is not clear, the reader can take a look at the book [JLR11]:

2.1 Asymptotic and set theory notation

We usually denote by $[n]$ the set of integers $\{1, \dots, n\}$. Given a set X, we denote the set of subsets of X of size k by $\binom{X}{k}$. The set of subsets of X (without considering the size) is written 2^X. We denote the *cardinality* (that is the number of elements in the set) of a set X as $|X|$. Let us now introduce some notation on asymptotic behaviour of functions. Consider $f(n), g(n)$ two functions. We compare the behaviour of this two functions when n tends to infinity. We say that $f = O(g)$ if there exists constants C, n_0 such that $|f(n)| \leq Cg(n)$. We say that $f = o(g)$ if $\lim_{n\to\infty} \frac{f}{g} \to 0$, i.e. for every $\varepsilon > 0$ there exists $n_0 = N_0(\varepsilon)$ such that $|f(n)| < \varepsilon g(n)$ for $n \geq n_0(\varepsilon)$.

2.2 Graph Theory

Graphs are one of the most fundamental structures in discrete mathematics and combinatorics. Recall that a *graph* G is a pair (V_G, E_G), where V_G is a set and $E_G \subset \binom{V_G}{2}$. V_G and E_G are the set of *vertices* and *edges* of G. We respectively denote by $|V_G|$, or equivalently $|G|$, the number of vertices of G and e_G the number of edges of G. The edge density of the graph G is defined as $\rho(G) = \frac{e_G}{|G|}$. In all this work we consider only simple graphs: neither loops nor multiple edges are allowed.

An edge joining the vertices u, v is denoted by uv, and we say that u and v are adjacent. Equivalently, the edge uv is incident to the vertices u and v. The degree of a single vertex v, denoted by $g(v)$, is the number of vertices adjacent to v. Given a subset of vertices $U \subset V$, the neighborhood of U is denoted by $N_G(U)$, and is composed by all vertices in G adjacent to U. Given G, a subgraph $H \subset G$ is any graph H with $V_H \subset V_G$ and $E_H \subset E_G$. Similarly, an induced subgraph H is a subgraph with $V_H \subset V_G$ and all the possible edges between vertices of V_H. A *path* in a graph G is succession of vertices $u_0, \dots, u_k$ such that $u_i u_{i+1} \in E_G$ for all $0 \leq i \leq k-1$, and $u_i \neq u_j$ for all i, j. A path where $u_0 = u_k$ is called a *cycle*. Given a subgraph $H \subset G$, we say that H is *connected* if for all $u, v \in H$ exists a path joining them. A subgraph $H \subset G$ is maximal if adding more vertices and edges from G makes H non connected. Indeed, if H is maximal we say that H is a *connected component* of G. We say G is a *connected graph* if it has only one connected component.

Let us introduce now some graphs with special properties. We denote by K_n the complete graph of n vertices. It is a graph which has all the possible $\binom{n}{2}$ edges joining two vertices. A complete graph is also usually called *clique*.
We denote by C_n the graph cycle of n vertices. It is a graph with n edges which consists on a cycle. Some more special graphs are well-known and widely used in graph theory, like the bipartite graph or the wheel graph, but in this work they are not mentioned anymore.

Finally, let us define a few graph parameters useful to describe a graph. The *girth* of a graph G, denoted by $g(G)$ is the size of its shortest cycle. $\alpha(G)$ is the size of the largest independent set (i.e. the largest set of vertices without any edge joining them) in G, and $\chi(G)$ denotes the chromatic number of G. That is the minimum number of colors need to perform a coloration of the vertices of G such that every two vertices joint by an edge have different colors. The minimum degree of a graph G is denoted by $\delta(G)$, and its maximum degree is denoted by $\Delta(G)$.

2.3 Probability Theory

A probability space is a measure space $(\Omega, \mathbb{A}, \rho)$, where Ω is the sample space (the set of all possible outcomes), $\mathbb{A}$ is a σ-algebra (the set of possible events) and ρ is a measure (the probability assigned to every event in $\mathbb{A}$). Given an event $A \subset \mathbb{A}$, the probability of A to happen is denoted as $\mathbb{P}(A)$. Given two events A, B, the conditional probability of A given B is defined as $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$. It can be interpreted as the probability of the event A to happen if the event B has already happened.

A random variable X is a measurable function $X : \Omega \to \mathbb{E}$, where Ω is the sample space of a probability space and $\mathbb{E}$ is usually $\mathbb{R}$, so given an event of the probability space X provides a real number. The event that a random variable takes values in $B \subseteq \mathbb{R}$ is denoted by $\mathbb{P}(X \in B)$. If the probability of an event to happen depends on a parameter n, we say that it occurs *with high probability*, from now referred to as *whp*, if its probability tends to 1 when n tends to infinity. In another words, making n big enough the probability is as close to 1 as desired. Given a sequence (finite or infinite) of random variables $X_1, \dots, X_n, \dots$, we say that they are independent and identically distributed (denoted as *i.i.d*) if they are independent and follow the same distribution X.

The expected value and variance of X are denoted by $\mathbb{E}[X]$ and $\mathbb{V}\mathrm{ar}[X]$, respectively. The covariance between two random variables X and Y is denoted by $\mathrm{Cov}[X, Y]$. We present several well-known probability inequalities, since they are basic to prove many results. Let us start with *Markov inequality*: let a be a real positive number and X be a positive random

variable (namely, $X \geq 0$). Then,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Let us consider a random variable X which is the sum of n i.i.d. Bernoulli random variables. Hence, Applying Markov inequality to the random variable e^{tX} and denoting $\mathbb{E}[X] = \mu$, we get the *Chernoff bounds*:

$$\mathbb{P}(X \geq (1+\delta)\mu) \leq e^{-\frac{\delta^2}{2+\delta}\mu} \text{ for all } \delta > 0;$$

$$\mathbb{P}(X \leq (1-\delta)\mu) \leq e^{-\frac{\delta^2}{2}\mu} \text{ for all } 0 < \delta < 1.$$

Another basic result in probability theory is *Chebyshev's inequality*. It states, in its standard form, that if X is a random variable with $\mathbb{E}[x] = \mu < \infty$, and $\mathbb{V}\mathrm{ar}[X] = \sigma^2 < \infty$, then for any $k > 0$,

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

The *Bonferroni inequalities* gives different lower and upper bounds for the probability of the union of some events. They state that, given $A_1, \ldots, A_n$ events in a probability space, $I \subset [r]$, $A_I = \cap_{i \in I} A_i$, $S_k = \sum_{|I|=k} \mathbb{P}(A_I)$, then:

$$t \text{ even} \implies \mathbb{P}(\cup A_i) \geq \sum_{i=1}^{t} (-1)^{i+1} S_i,$$

$$t \text{ odd} \implies \mathbb{P}(\cup A_i) \leq \sum_{i=1}^{t} (-1)^{i+1} S_i.$$

The *law of total probabilities* is also essential for some proofs. For any partition of the probability space $\Omega = \bigcup \omega_i$,

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X|\omega_i]\mathbb{P}(\omega_i).$$

In particular, taking $X = 1_\omega$ the indicator function for all $\omega \subset \Omega$, we obtain

$$\mathbb{P}(\omega) = \sum_i \mathbb{P}[\omega|\omega_i]\mathbb{P}(\omega_i).$$

To finish this chapter, let us recall some properties of the most common random variables. First of all, a random variable X follows a Bernoulli distribution of parameter p ($X \sim Bern(p)$) if it takes value 1 with probability p, and value 0 with probability $q = 1 - p$. Its mean is p and its variance is pq. The sum of n Bernoulli random variables is called a Binomial distribution $X \sim Bin(n, p)$. Its mean is np and its variance is npq. Another important random variable model is Poisson. A random variable X follows a Poisson distribution of paramater λ if $\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$. We denote this random variable as $X \sim Po(\lambda)$. Its mean and its variance are both λ. Finally, the most common probability distribution is the normal or gaussian. We denote a random variable as $N(\mu, \sigma^2)$ if is normally distributed with mean μ and variance σ^2.

3. Random graph models

A random graph model is build by considering a set of labelled vertices and adding successive edges between them at random. We consider only finite sets of vertices in this work. If the set is infinite, the model is called an infinite random graph and it is another field to study with different results and applications. The way we add edges to the graph defines the random graph model. There are many random graph models, coming from different points of view (random regular graphs, random dot-product model...), but the most common (and the ones we study in this work) are Gilbert model (also known as $G(n,p)$) and Erdős-Rényi model (also known as $G(n,M)$).

The goal of studying random graphs usually is to understand what properties (containing a specific subgraph, connectivity and hamiltonicity are the most studied in this work) accomplishes the random graph as a function of some parameters that appears on its construction. More precisely, a main goal is to know at what time a particular graph property is likely to arise. This characterization for random graph models $G(n,p)$ and $G(n,M)$ is further studied in the next chapter, focusing in threshold phenomena for the appearance of some properties.

3.1 The origins of random graphs

We see a pair of examples to illustrate the first ideas related to random graphs. By this way we can start understanding its utility, since they are so representative of the kind of problems that can be faced with random graph theory.

First of all, we introduce the probabilistic method. This mathematical technique consists to build a probability space and prove that a certain event (we can understand it as a property that some elements of the sample space fulfill) has positive probability in this space. In this way, we can ensure that there exists an element in the sample space which accomplishes the desired property. Let us remark that it is a non-constructive method. In general, it does not provide an algorithm to find the solution. This is so useful in problems where getting explicit constructions is extremely difficult.

The example we show to introduce this method belongs to Ramsey theory. This mathematical field, developed basically by Frank Plumpton Ramsey at the 1920's, studies the appearance of order in things such as sequences, groups or specially graphs. Indeed, the example we show is a result of graph coloration in terms of the so-called Ramsey numbers, defined as consequence of the following theorem, which we not prove:

Theorem 3.1. Let $k \in \mathbb{N}$. Then there exists an $n \in \mathbb{N}$ such that any 2-colored K_n graph contains a monochromatic subgraph K_k of K_n.

Notice the importance of this theorem, since it provides an orderly substructure (the

monochromatic subgraph K_k), that must be included in a larger 2-colored structure. From this idea, the *Ramsey numbers* are defined as follows:

Definition 3.2. The *Ramsey number* $R(k, l)$ is the smallest integer such that in any two-coloring of the edges of a complete graph of n vertices K_n, it contains either a monochromatic subgraph K_k or a monochromatic subgraph K_l.

Notice that the number of edges of a K_n grows enormously with n. Because of this fact, it is so difficult to compute the exact value of the Ramsey number. However, the following proposition gives a lower bound for the *diagonal Ramsey number*, defined as the Ramsey number $R(k, l)$ with $k = l$. Notice that because of theorem 3.1, it exists and it is a finite number. The proof of this proposition is based on the probabilistic method, and it is a good example to see how this argument shows off:

Proposition 3.3. *[AS04]* If $\binom{n}{k}2^{1-\binom{n}{k}} < 1$, then $R(k, k) > n$.

Proof. The key idea is to color independently each edge. More precisely, let us take a random coloring (with two colors, let us say red and blue) of the edges K_n, with every edge colored red with probability p.
Let A_R the following event: The induced subgraph in a set R of k vertices is monochromatic (either red or blue). Thus,

$$\mathbb{P}(A_R) = \frac{1}{2^{\binom{k}{2}}} + \frac{1}{2^{\binom{k}{2}}} = 2^{1-\binom{k}{2}}.$$

Taking this result for all possible set R and using the union bound for probabilities we get that

$$\mathbb{P}(\cup A_R) \leq \sum 2^{1-\binom{k}{2}} = \binom{n}{k} \cdot 2^{1-\binom{k}{2}}.$$

We impose this value to be strictly smaller than 1. By this way, we can ensure that the complementary event has probability strictly positive. Here is where the probabilistic argument shows off, as it means that there exists a 2-coloration without any monochromatic K_k.
Thus, $R(k, k) > n$, as we desired to prove. □

This proof confirms that getting real bounds for the *diagonal Ramsey numbers* involve very large numbers. In fact, it is impossible to get the value of $R(k, k)$ for $k \geq 6$, as the number of cases to count increase in such a way that make them intractable. As an anecdote that exemplifies this fact, Erdős let in an article in the 90's the following quote: *"Suppose aliens invade the earth and threaten to obliterate it in a year's time unless human beings can find $R(5, 5)$. We could marshal the world's best minds and fastest computers, and within a year we could probably calculate the value. If the aliens demanded $R(6, 6)$, however, we would have no choice but to launch a preemptive attack."*

In spite of the difficulty of the problem, the probabilistic method allows to find a sufficiently good lower bound, showing its practical utility. To finish with diagonal Ramsey numbers, let us comment the best lower and upper bounds found nowadays. The classic upper bound, due to Erdős and Szekeres [ES35], states that

$$R(k,k) \leq \binom{k-1+l}{k-1}.$$

Some improvements to this upper bound have been found. Let us show the best upper bound, a pretty recent result (2009), shown in [Con09]. It gives the following result:

$$R(k,k) \leq (k-1)^{-C\frac{\log k-1}{\log\log k-1}} \binom{2(k-1)}{k-1}.$$

Lower bounds have been also widely studied, getting results either by probabilistic arguments or by constructive methods. An overview of this results can be found on [CG83], we only cite the classic lower bound, due again to Erdős: [Erd47]

$$R(k,k) > \frac{k2^{\frac{k}{2}}}{e\sqrt{2}}.$$

At this point, once the probabilistic method is already introduced, let us show the first appearance of the random graph concept. It is defined in a very simple way in order to use it in a proof. It is a proof based in the probabilistic method that illustrates well the utility of this kind of structure. After this theorem, stated by Erdős and Rényi in 1959, they realized that random graphs could be a very interesting field to study, and published the paper [ER59]. This paper showed to the mathematical community that random graphs are a very interesting field, and since that moment their study has given many results in very different areas. Before the proposition, we state a previous lemma which is necessary in the main proof.

Lemma 3.4. Let G be a graph on n vertices. Then $\alpha(G)\cdot\chi(G) \geq n$.

Proof. We consider a coloration of G. By definition of chromatic number, there are $\chi(G)$ subsets, and every color class is an independent set. The sum of the sizes of every subset is n, and every subset has size smaller or equal than $\alpha(G)$, so this fact gives the following inequality:

$$\alpha(G)\cdot\chi(G) \geq n,$$

as we wanted to prove. □

The previous lemma is so useful in the proof of the theorem, since it allows us to work with $\alpha(G)$ instead of using the chromatic number, which is very difficult to control. Indeed, thanks to the lemma to find an upper bound of $\chi(G)$ it is equivalent to find a lower bound for $\alpha(G)$, and this last option is much easier. The theorem itself provides a surprising result, since by intuition a graph with high chromatic number has in general many edges, and a graph with high girth has in general few edges, but the theorem states that there exists a graph G with both $\alpha(G)$ and $\chi(G)$ arbitrarily large.

Theorem 3.5. *[Erd59]* For any choice of k and l positive numbers there exists a graph G with $g(G) > l$ and $\chi(G) > k$.

Proof. Let $X_{\leq l}$ be the number of cycles of length smaller than l in a graph G. Taking G any random graph, $X_{\leq l}$ is a random variable. Then, we can compute its expected value. For any possible length of the cycle, we choose its vertices, taking into account the symmetry of the vertices and the rotation of the cycle, and we multiply by the probability of every edge to exist, we consider it is a constant p. Hence, we obtain:

$$\mathbb{E}[X_{\leq l}] = \sum_{i=3}^{l} \frac{n(n-1)...(n-i+1)}{2i} \cdot p^i.$$

We use now that $n(n-1)...(n-i+1) \leq n^i$. We also choose $p = n^{\theta-1}$, taking $\theta < \frac{1}{l}$. Substituting in the previous equation, we get

$$\mathbb{E}[X_{\leq l}] \leq \sum_{i=3}^{l} \frac{n^{\theta i}}{2i} \leq \sum_{i=3}^{l} \frac{n^{1-\varepsilon i}}{2i},$$

so $E[X_{\leq l}]$ is $o(n)$ by definition. Using Markov inequality,

$$\mathbb{P}\left(X_{\leq l} \geq \frac{n}{2}\right) \leq \frac{E[X_{\leq l}]}{\frac{n}{2}}.$$

Thus, $\frac{\mathbb{E}[X_{\leq l}]}{\frac{n}{2}} = 2\frac{\mathbb{E}[X_{\leq l}]}{n} = o(1)$, so $\mathbb{P}\left(X \geq \frac{n}{2}\right) = o(1)$.
At this point, we use lemma 3.4 to work with $\alpha(G)$ instead of $\chi(G)$. By the union bound and using that $(1-p)^n \leq e^{-pn}$, we obtain

$$\mathbb{P}(\alpha \geq x) \leq \binom{n}{x}(1-p)^{\binom{x}{2}} < \left(ne^{\frac{-p(x-1)}{2}}\right)^x.$$

Considering $x = \frac{3}{p}\log n$, then $\mathbb{P}(\alpha \geq x) = o(1)$.

As both events are o(1), let us take n large enough such that $\mathbb{P}(\alpha \geq x) < \frac{1}{2}$ and such that $\mathbb{P}\left(X_{\leq l} \geq \frac{n}{2}\right) < \frac{1}{2}$. So it exists a graph G such that $\alpha(G) < 3n^{1-\theta}\log n$ and it has at least $\frac{n}{2}$ cycles of length smaller or equal than l. For every one of this cycles, we subtract a vertex. We obtain the graph G^* with at least $\frac{n}{2}$ vertices and all cycles with length greater than l (by definition $g(G^*) > l$).
A maximal independent set in G, with size $\alpha(G)$, is maximal also in G^*, so $\alpha(G^*) \leq \alpha(G)$.
Using this two last inequalities,

$$\chi(G^*) \geq \frac{|G^*|}{\alpha(G^*)} \geq \frac{\frac{n}{2}}{\alpha(G)} \geq \frac{\frac{n}{2}}{3n^{1-\theta}\log n} = \frac{n^\theta}{6\log n}.$$

At this point, we can take this last expression greater enough such that the graph G^* has girth greater than l and chromatic number greater than k, as we wanted to show. □

The proof of the proposition shows that this kind of random graph model is a very useful tool when dealing with graph properties. That is because it is a concept very prone to use in a probabilistic argument, which is very common in this type of proofs. By this way, using them as an auxiliary element in the proof, this concept appeared and started to be studied in the forthcoming years.

3.2 Definition of $G(n,p)$ and $G(n,M)$

As we mentioned in the introduction, we refer to the Erdős-Rényi model for random graphs. Let us define the two basic models, the binomial and the uniform random graph model. The first one comes from a probabilistic point of view, while the second one comes from combinatorics theory.

The idea behind the notion of binomial random graph, which we denote by $G(n,p)$, is to consider a set of n labelled vertices, and for every pair of edges we flip a coin, with probability p of success. We call the probability of failure $q = 1-p$. Only if we have success, we draw the edge which joins the two vertices. Let us formalize this definition, as can be found in [Die05]. Let V be a fixed set of n elements, say $V = [n]$. For every potential edge $e \in [V]^2$, we define a sample space $\Omega_e = \{0_e, 1_e\}$, where 0_e means that the edge e does not exist, and 1_e mean that it exists. We choose $\mathbb{P}(1_e) = p$, $\mathbb{P}(0_e) = q$, so we define the probability space Ω by taking the product of all probability samples. An element in Ω is a graph G on V with edge set $E(G) = \{e|\omega(e) = 1_e\}$, where ω is a map which assign a certain probability P_e to every edge e. Finally, any random graph G is an event in this probability space Ω. Considering a uniform map ω, i.e. it assigns the same probability p to all edges, the random graph we get follows a $G(n,p)$ model. By definition, all the events are independent, since the existence of a fixed edge is independent of the existence of the rest of possible edges.

We are not going to state a formal definition of $G(n,M)$ as we did for $G(n,p)$, but we give the reader the main ideas. The uniform random graph rises from enumerative principles. We choose one graph uniformly at random among all graphs on M edges and n vertices. Hence, the probability of a graph G is

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

The two models are very similar, in terms that we precise later, but in order to study them sometimes it is much easier to work with one of both instead than work with the other. In particular, we prove the majority of results for $G(n,p)$, but taking into account that they are also valid for $G(n,M)$ if some conditions are fulfilled.

3.3 Graph properties. Asymptotic equivalence

Let us first set the definition of a graph property. A subfamily of graphs Q is a graph property if $Q \subset \mathbb{G}$, where $\mathbb{G}$ is the family of all graphs in a probability space. If G belongs to the subfamily of random graphs Q, we write $G \in Q$, and we say that G fulfills the property Q. Graph properties are widely studied because the graphs belonging to a given subfamily often can be characterized thanks to the study of the property. Some well-known examples of graph properties are planarity, connectivity, hamiltonicity or existence of eulerian cycles. Notice that we are considering this properties as subfamilies of graphs, so for us a connected graph is a graph such that it belongs to the subfamily of connected graphs, and that is by definition the subset of all graphs that consists on only one connected component. Let us first introduce the concepts of monotone, increasing and decreasing properties:

Definition 3.6. [JLR11] Let Q be a graph property. Then:

- Q is *increasing* if for all A, B subsets of random graphs, $A \subset B$, $A \in Q$, then $B \in Q$.
- Q is *decreasing* if for all A, B subsets of random graphs, $A \subset B$, $B \in Q$ then $A \in Q$.
- If a property fulfills one of the previous two definitions, we say it is a monotone property.

First, we prove that asymptotically if $p\binom{n}{2}$ is close to M the properties fulfilled by $G(n, M)$ are also fulfilled by $G(n, p)$. To show the reciprocal implication, we need the property to be monotone. This is not a big problem, as many of the most important graph properties are monotone. We first show a lemma which is necessary in the second proof.

Lemma 3.7. *[JLR11]* Let Q be an increasing property. Assume that $0 \leq p_1 \leq p_2 \leq 1$ and $0 \leq M_1 \leq M_2 \leq N$. Then $\mathbb{P}(G(n, p_1) \in Q) \leq \mathbb{P}(G(n, p_2) \in Q)$ and $\mathbb{P}(G(n, M_1) \in Q) \leq \mathbb{P}(G(n, M_2) \in Q)$.

Proof. Observe first that the lemma is obvious for the $G(n, M)$ model: if $M_1 \leq M_2$ then $G(n, M_1) \subset G(n, M_2)$, because every graph of M_1 edges can be expanded to graph on M_2 edges by adding $M_2 - M_1$ edges. So the sample space $G(n, M_1)$ must be included on $G(n, M_2)$. Since Q is increasing, $G(n, M_1) \in Q$ implies that $G(n, M_2) \in Q$. Hence, we have

$$\mathbb{P}(G(n, M_1) \in Q) \leq \mathbb{P}(G(n, M_2) \in Q).$$

To prove the result for $G(n, p)$, the idea is to split the edges of $G(n, p_2)$ into two subsets. We consider $p_0 = \frac{p_2 - p_1}{1 - p_1}$. Observe that $p_2 = p_0 + p_1 - p_0 p_1$, where substracting the term $p_0 p_1$ corresponds to removing the common edges. Thus, we can write that $G(n, p_2) = G(n, p_0) \cup G(n, p_1)$. Immediately we deduce that $G(n, p_1) \subset G(n, p_2)$, and as Q is increasing, we have

$$\mathbb{P}(G(n, p_1) \in Q) \leq \mathbb{P}(G(n, p_2) \in Q),$$

as we wanted to prove. □

The number of edges in $G(n,p)$ follows a binomial distribution, since we sum over all the $\binom{n}{2}$ Bernoulli random variables that correspond to any possible edges. Hence, the expectation for the number of edges of $G(n,p)$ is $\binom{n}{2}p$, so one may expect that both $G(n,p)$ and $G(n,M)$ are similar when M is close to $\binom{n}{2}p$. This intuitive fact is formalized in the following propositions. From now we define $N = \binom{n}{2}$. Let us first prove that if $G(n,M)$ fulfills a certain property and M is close to Np, then the property is also fulfilled for $G(n,p)$:

Proposition 3.8. *[JLR11]* Let Q be a property, $p = p(n)$, $0 \leq a \leq 1$, and $M = M(n)$ such that $M = Np + O(\sqrt{Npq})$. Then,

$$\lim_{n\to\infty} \mathbb{P}(G(n,M) \in Q) \to a \implies \lim_{n\to\infty} \mathbb{P}(G(n,p) \in Q) \to a.$$

Proof. Let C be a constant. We define $M(C)$ as the subset of all values of M that belong to the closed ball of center Np and radius $C\sqrt{Npq}$. Let M^* be the element of $M(C)$ that minimizes $\mathbb{P}(G(n,M) \in Q)$. It exists since $M(C)$ is a compact (that is why we choose the closed ball) and that probability is a bounded function, because by definition its value must be in $[0,1]$. By the law of total probability,

$$\mathbb{P}(G(n,p) \in Q) = \sum_{M=0}^{N} \mathbb{P}(G(n,p) \in Q | e_{G(n,p)} = M) = \sum_{M=0}^{N} \mathbb{P}(G(n,M) \in Q | e_{G(n,p)} = M) \geq$$

$$\geq \sum_{M \in M(C)} \mathbb{P}(G(n,M^*) \in Q | e_{G(n,p)} = M) = \mathbb{P}(G(n,M^*) \in Q)\mathbb{P}(e_{G(n,p)} \in M(C)).$$

In the second inequality we have used that $M(C)$ is a subset of all possible values of M and furthermore M^* minimizes the probability by definition. In the last equality we use the definition of conditional probability for every term in the sum, and we add all the terms.
We can assume that $\mathbb{P}(G(n,M^*) \in Q) \to a$ because $M^* \in M(C)$, so by definition $M^* = Np + O(\sqrt{Npq})$. By Chebyshev inequality, since the variance of a binomial distribution is Npq, we have that

$$\mathbb{P}(e_{G(n,p)} \notin M(C)) \leq \frac{\mathbb{V}\mathrm{ar}\,[G(n,p)]}{(C\sqrt{Npq})^2} = \frac{1}{C^2}.$$

Thus, $\mathbb{P}(e_{G(n,p)} \in M(C)) \geq 1 - \frac{1}{C^2}$. Going back to the previous equation, $\liminf \mathbb{P}(G(n,p) \in Q) \geq a(1 - \frac{1}{C^2})$. Let us consider now M^{**} which maximizes $\mathbb{P}(G(n,M) \in Q)$ for all $M \in M(C)$, using analogous arguments as we did with M^* we obtain

$$\mathbb{P}(G(n,p) \in Q \leq \mathbb{P}(G(n,M^{**}) \in Q) + \mathbb{P}(e_{G(n,p)} \notin M(C)) \leq \mathbb{P}(G(n,M^{**}) \in Q + \frac{1}{C^2}.$$

Hence, we obtain that $\limsup \mathbb{P}(G(n,p) \in Q \leq a + \frac{1}{C^2}$. So at this point we have $\liminf \mathbb{P}(G(n,p) \in Q \geq a\left(1 - \frac{1}{C^2}\right)$, and $\limsup \mathbb{P}(G(n,p) \in Q) \leq a\left(1 + \frac{1}{C^2}\right)$. As C is arbitrary, letting n tend to infinity and taking C arbitrary large we obtain

$$\lim \mathbb{P}(G(n,p) \in Q) \to a,$$

as we desired to prove. □

We have proven only one implication, now if we are able to prove the reciprocal one it is immediate that properties for both models are equivalent. However, an extra condition is necessary, that is that the given property has to be monotone. This is not a big problem, as the graph properties we study (connectivity, hamiltonicity...) are indeed monotone. The statement for this proposition is the following:

Proposition 3.9. *[JLR11]* Let Q be a monotone property, $0 \leq M \leq N$, $0 \leq a \leq 1$, and $p = p(n)$ such that $p = \frac{M}{N} + O\left(\sqrt{\frac{M(N-M)}{N^3}}\right)$. Then,

$$\lim_{n\to\infty} \mathbb{P}(G(n,p) \in Q) \to a \implies \lim_{n\to\infty} \mathbb{P}(G(n,M) \in Q) \to a.$$

Proof. Let us consider, without loss of generality, the case where Q is increasing. Let C be a constant, and let us define $p_0 = \frac{M}{N}, q_0 = 1 - p_0, p_+ = \min(p_0 + C\sqrt{\frac{p_0 q_0}{N}}, 1)$, and $p_- = \max(p_0 - C\sqrt{\frac{p_0 q_0}{N}}, 0)$. Using the same arguments as in the previous proof of proposition 3.8, we see that

$$\mathbb{P}(G(n,p_+) \in Q) \geq \sum_{M' \geq M} \mathbb{P}(G(n,M'))\mathbb{P}(|G(n,p_+)| = M').$$

We can bound even more this value. Using this last expression, we obtain:

$$\mathbb{P}(G(n,p_+) \in Q) \geq \mathbb{P}(G(n,M) \in Q) - \mathbb{P}(|G(n,p_+)| < M).$$

Similarly,

$$\mathbb{P}(G(n,p_-) \in Q \leq \mathbb{P}(G(n,M) \in Q) + \mathbb{P}(|G(n,p_-)| > M).$$

We assume $1 \leq M \leq N-1$ (as both cases $M = 0, M = N$ are trivial), and let us define $\delta(C) = C^{-2} + \sqrt{2}C^{-1}$. Thus, by the Chebyshev inequality,

$$\mathbb{P}(|G(n,p_-)| > M) \leq \frac{Np_-(1-p_-)}{(Np_0 - Np_-)^2} \leq \frac{Np_0q_0 + C\sqrt{Np_0q_0}}{C^2Np_0q_0} \leq \frac{1}{C^2} + \frac{1}{C\sqrt{Np_0q_0}} \leq \delta(C).$$

In the last inequality, we apply that $Np_0q_0 = \frac{M(N-M)}{N} \leq 2$. Then, similarly $\mathbb{P}(|G(n,p_+)| < M) < \delta(C)$. Since $\lim_{n\to\infty} \mathbb{P}(G(n,p_+) \in Q) = \lim_{n\to\infty} \mathbb{P}(G(n,p_-) \in Q) \to a$ by assumption, and using the inequalities we have proven, we obtain

$$a - \delta(C) \leq \liminf_{n\to\infty} \mathbb{P}(G(n,M) \in Q) \leq \limsup_{n\to\infty} \mathbb{P}(G(n,M) \in Q) \leq a + \delta(C).$$

Thus, the result follows by letting C be arbitrarily large, since it implies that $\delta(C)$ tends to 0, and we conclude that $\lim_{n\to\infty} \mathbb{P}(G(n,M) \in Q) = a$, as we desired to prove. □

In the following chapter, we study what is the behaviour of $G(n,p)$ respect to some well-known properties, when p varies. This is translated to the fact that the structure of the graph dramatically changes once p rises from 0 to 1. Usually, the transition probability is very abrupt, and the graph experiments heavy changes. The study of this changes is a central topic in this work, as it give us an overview of the general shape of $G(n,p)$ for a given p.

4. Properties of $G(n,p)$: Threshold functions and subgraphs

In this chapter, we introduce a central concept in this work, that is threshold functions. Since we consider $p = p(n)$, it arises as a natural question how is the evolution of the graph when p rises from 0 to 1. This behaviour is characterized thanks to threshold functions. Informally speaking, a threshold function for a monotone property is an asymptotic family of functions, depending on n, such that if we take p smaller than the threshold (asymptotically) the random graph $G(n,p)$ does not accomplish the desired property *whp*. On the other hand, if we take p greater than the threshold, then $G(n,p)$ accomplishes the desired property *whp*. The same definition can be set for $G(n,M)$, since we have seen in the previous chapter that for monotone properties both models have the same asymptotic behaviour. Once we set the definition and the existence of threshold functions, in this chapter we work threshold functions for the property *"To contain a concrete subgraph"* for some important subgraphs. Let us first state the formal definition of threshold functions for both models.

Definition 4.1. Given an increasing property Q (the definition is analogue for a decreasing property), the following holds:

- A sequence $P = P(n)$ is a threshold function if $\mathbb{P}(G(n,p) \in Q) \to 0$, if $p = o(P)$ and $\mathbb{P}(G(n,p) \in Q) \to 1$, if $p = O(P)$.
- A sequence $M = M(n)$ is a threshold function if $\mathbb{P}(G(n,M^*) \in Q) \to 0$, if $M^* = o(M)$ and $\mathbb{P}(G(n,M^*) \in Q) \to 1$, if $M^* = O(M)$.

Notice that given a threshold function for a monotone property we can clearly characterize this property, as we know for what values of p it is accomplished and for what values of p it is not. The only question which remains to study is what happens when p is of the same asymptotic order than the threshold, that is equivalent to study how the property arises. This case must be further studied for every property, since it is a more difficult question and it is not possible to get general results. The question now is if any monotone property can be characterized using threshold functions. The answer is yes, and that is the reason why they are an extremely useful tool in this field. The following theorem, due to Bollobás and Thomasson, gives the desired result:

Theorem 4.2 ([JLR11])**.** Any monotone property has a threshold.

Proof. We prove the result for Q increasing property (the very same argument applies when Q is decreasing). Let $0 < \varepsilon < 1$. Let m be an integer such that $(1-\varepsilon)^m \leq \varepsilon$ (observe that such m exists). We consider now m independent copies of $G(n,p(\varepsilon))$, which we denote by $G(n,p_1),...,G(n,p_m)$, respectively. All these copies are taken over the same set of vertices $[n]$. It is possible that some of them share edges.

We denote by $G(n, p')$ the union of all of them, for a certain value of p'. In fact, we can easily deduce p'. The probability of an edge to be in $G(n, p')$ is 1 minus the probability of not being in any graph, and every copy is independent, namely

$$p' = 1 - (1 - p(\varepsilon))^m.$$

By Lemma 3.7, as $1 - (1 - p(\varepsilon))^m \leq mp(\varepsilon)$, we have

$$\mathbb{P}\left(\cup_{i=1}^{m} G(n, p_i) \in Q\right) \leq \mathbb{P}(G(n, mp(\varepsilon)) \in Q).$$

Also, $\mathbb{P}(\cup G(n, p_i) \notin Q) \leq \mathbb{P}(G(n, p_i) \notin Q \text{ for all i}) = (1 - \varepsilon)^m \leq \varepsilon$, because of the choice of m. So, we have that

$$\mathbb{P}(G(n, p(\varepsilon)) \in Q) \geq \mathbb{P}(\cup G(n, p_i) \in Q) \geq 1 - \varepsilon.$$

Thus, as $p(1 - \varepsilon) \leq mp(\varepsilon)$, taking $\varepsilon \in (0, \frac{1}{2})$ we have $p(\varepsilon) \leq p(\frac{1}{2}) \leq p(1 - \varepsilon) \leq mp(\varepsilon)$. At this point, m depends on ε but not on n, so $p(\varepsilon), p(\frac{1}{2})$ and $p(1 - \varepsilon)$ are of the same asymptotic order. It can be proved (it is in [JLR11]) that in this case $p(\frac{1}{2})$ is a threshold for Q. Notice that the key fact is that we define $\varepsilon \in (0, \frac{1}{2})$, so it implies $\varepsilon < \frac{1}{2} < 1 - \varepsilon$, and it concludes the proof. □

Let us start by looking at the threshold for some important monotone properties in a graph, once we have ensured that it exists. We start with the property *"To contain a a certain subgraph"*. In fact, some graph properties can be studied as the property of having a concrete subgraph. Let us start by proving the threshold when the subgraph is K_4. Then, we generalize the proof to K_r for any fixed r. We also study what happens when the subgraph we look for is a more general random graph, as a balanced subgraph. Following this order, we get an idea of the arguments used in a concrete case (the K_4) and then we continue getting more general results that use the same concepts.

Before the proof, we state a useful inequality. Thus, we define $\Delta = \sum_{i \sim j} P(A_i \cap A_j)$, where i, j are sets and A_i, A_j are given events. The following result states that the random variable X is around $\mathbb{E}[X]$ if a certain condition for Δ is fulfilled:

Proposition 4.3. *[AS04]* If $\mathbb{E}[X] \to \infty$ and $\Delta = o(\mathbb{E}[X]^2)$, then $X > 0$ *whp*. Furthermore, $X \sim \mathbb{E}[X]$ *whp*.

Proof. We denote $\mathbb{E}[X] = \mu, \mathbb{V}\mathrm{ar}[X] = \sigma^2$ and $\lambda = \frac{\mu}{\sigma}$. Then, using Chebyshev inequality

$$\mathbb{P}(X = 0) \leq \mathbb{P}(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2} = \frac{\sigma^2}{\mu^2}.$$

Hence, if $\mathbb{V}\mathrm{ar}[X] = o(\mathbb{E}[X]^2)$ then $\mathbb{P}(X = 0) = 0$, so $X > 0$ *whp*. Moreover, for any $\varepsilon > 0$, we have that

$$\mathbb{P}(|X - \mu| \geq \varepsilon\mu) \leq \frac{\sigma^2}{\varepsilon^2\mu^2}.$$

Hence, if $\mathbb{V}\mathrm{ar}[X] = o(\mathbb{E}[X]^2)$ then $X \sim \mathbb{E}[X]$ *whp*. Consider $X = X_1 + \cdots + X_m$, where X_i is the indicator random variable for the corresponding event A_i. Thus, if $i \sim j$, $\mathrm{Cov}[X_i, X_j] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \leq \mathbb{E}[X_i X_j] = \mathbb{P}(A_i \cap A_j)$.If $i \not\sim j$, then A_i and A_j are independent, so $\mathrm{Cov}[X_i, X_j] = 0$. In any case, $\mathbb{V}\mathrm{ar}[X] \leq \mathbb{E}[X] + \Delta$. If $\mathbb{E}[X] \to \infty$ and $\Delta = o(\mathbb{E}[X]^2)$, then $\mathbb{V}\mathrm{ar}[X] = o(\mathbb{E}[X]^2)$. Hence, joining this result with the previous inequalities, we get the desired result. □

Let us now state and prove the first explicit threshold function we see in this work. The property we look for is *"To contain a K_r for any fixed r"*. This is an important property since the complete graph has been widely studied and its behaviour is well-known. Indeed, the study of this graph property in all its variants (finding the maximum r such that K_r is contained, testing if a graph contains a K_r for r greater than a given size, finding the maximal K_r in the graph,...) is a big field of research. It is generally referred to as the *Clique problem*. It has applications on social networks, computational science, biology, etc. For any graph the property we state is a very difficult problem, called the *clique decision problem*. However, the special structure of random graphs, focusing on the existence of threshold function for monotone properties, allows us to achieve pretty good results on this problem. We start with $r = 4$ (a representation for K_4 is shown in figure 1) to better understand the concepts. Then, we generalize the theorem for every r.

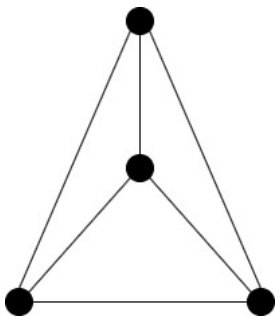


Figure 1: K_4 representation

Theorem 4.4. *[AS04]* The property *"To contain a K_4"* for any random graph $G(n,p)$ has threshold function $p = n^{\frac{-2}{3}}$.

Proof. Let X be the number of K_4 contained in G. If S is a set of four vertices in G, we consider the indicator random variable X_S for S to be a K_4, and by this way

$$X = \sum_{|S|=4} X_S.$$

As all edges must appear, each one with probability p, clearly $\mathbb{E}[X_S] = p^6$. By linearity of the

expectation, we can compute easily

$$\mathbb{E}[X] = \sum_{|S|=4} X_S = \binom{n}{4} p^6 \approx n^4 p^6.$$

Taking $p(n) = o\left(n^{\frac{-2}{3}}\right)$, then $\mathbb{E}[X] \approx n^4 n^{-4-\varepsilon} = n^{-\varepsilon} \to 0$, as $\varepsilon > 0$ and n tending to infinity. We have $X \geq 0$ and $\mathbb{E}[X] = 0$, so X must be 0 *whp*. Therefore, for $p(n) = o\left(n^{\frac{-2}{3}}\right)$ there is not any K_4 contained in G with probability close to 1.In the case of taking $p(n) = O\left(n^{\frac{-2}{3}}\right)$, clearly $\mathbb{E}[X] \to \infty$, but it does not imply that $X > 0$ directly. We have seen that the expected value for K_4 is $n^4 p^6$, so $\mathbb{E}[X]^2 = n^8 p^{12}$. Let us compute the expected number of intersection between different K_4 into G. Notice that the only possible intersections are sharing one or three edges. First, if the intersection is one edge, we are in the situation shown in figure 2:

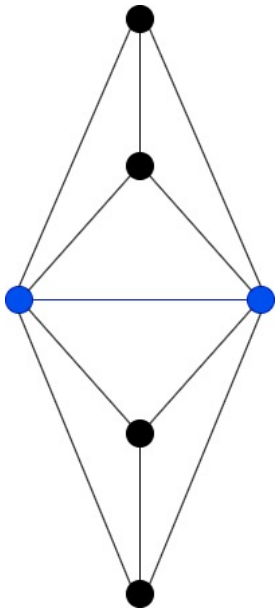


Figure 2: Two copies of K_4 sharing one edge and two vertices, painted in blue

We fix this edge (asymptotically $\approx n^2 p$) and we choose the rest of vertices and edges to complete the two K_4:

$$\mathbb{E}[X] = n^2 p (n^2 p^3)^2 = n^6 p^{11}.$$

This is $o(\mathbb{E}[X]^2)$, since

$$\lim_{n\to\infty} \frac{n^6 p^{11}}{n^8 p^{12}} = \lim_{n\to\infty} \frac{1}{n^2 p} \to 0.$$

Now, if the intersection consists of three edges, we are in the situation shown in figure 3:

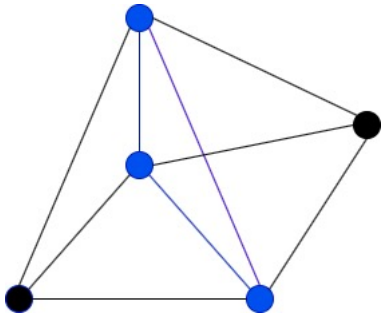


Figure 3: Two copies of K_4 sharing three edges and three vertices, painted in blue

Using the same arguments to count asymptotically the expectation of this kind of intersections, we get as expected value

$$\mathbb{E}[X] = n^3 p^3 (np^3)^2 = n^5 p^9.$$

Computing the limit as before, this is $o(\mathbb{E}[X]^2)$. So, since both are $o(\mathbb{E}[X]^2)$, the sum Δ is also $o(\mathbb{E}[X]^2)$. Because of proposition 4.3, it implies that $X > 0$ *whp*, so the probability of having a K_4 as a subgraph of G tends to 1 for $p(n) = O\left(n^{\frac{-2}{3}}\right)$. Therefore, $p(n) = n^{\frac{-2}{3}}$ is a threshold, as we desired to show. □

Notice that this threshold function comes in a natural way with this proof when computing $\mathbb{E}[X] \approx n^4 p^6$. We built it imposing that $\mathbb{E}[X] = o(1)$. This expected value can be generalized for any $k > 4$ as $n^k p^{\binom{k}{2}}$, since K_k has k vertices and $\binom{k}{2}$ edges. Therefore, we deduce the following theorem:

Theorem 4.5. *[AS04]* The property *"To contain a K_k"* for any k and for any random graph $G(n,p)$ has threshold function $p = n^{\frac{-v}{e}}$, *where* $v = k$, $e = \binom{k}{2}$.

Proof. The proof is quite similar to the one of theorem 4.4. Let X be the number of K_k contained in G. If S is a set of k vertices in G, we consider the indicator random variable X_S for S to be a K_k. Calculating the expectation as we did before,

$$\mathbb{E}[X] = \sum_{|S|=k} X_S = \binom{n}{k} p^{\binom{k}{2}} \approx n^k p^{\binom{k}{2}}.$$

Taking $v = k$, $e = \binom{k}{2}$ and $p(n) = o\left(n^{\frac{-v}{e}}\right)$, then $\mathbb{E}[X] \approx n^v n^{-v-\varepsilon} = n^{-\varepsilon} \to 0$, as $\varepsilon > 0$ and $n \to \infty$. We have $X \geq 0$ and $\mathbb{E}[X] = 0$, so X must be 0 *whp*. Therefore, for $p(n) = o\left(n^{\frac{-v}{e}}\right)$ there is not any K_k contained in G *whp*. As happened in theorem 4.4, in the case of taking $p(n)O\left(n^{\frac{-v}{e}}\right)$, clearly $\mathbb{E}[X] \to \infty$, but it does not imply that $X > 0$. We have seen that the expected value for K_k is $n^k p^e$, so $\mathbb{E}[X]^2 = n^{2k} p^{2e}$. Let us compute the expected number

of intersection between different K_k into G (considering intersections that share l edges, for $l = 2, \dots, l-1$).For every l, if the intersection are l vertices we fix them and all the edges joining them (asymptotically $\approx n^l p^{\binom{l}{2}}$) and we choose the rest of vertices and edges to complete the two K_k:

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

This is $o(\mathbb{E}[X]^2)$. The limit to check it can be computed easily as before, it remains

$$\lim_{n\to\infty} n^{-l} p^{-\binom{l}{2}} \to 0.$$

Since this is true for all l, the sum Δ is also $o(\mathbb{E}[X]^2)$. Because of 4.3, it implies that $X > 0$ *whp*, so the probability of having a K_k as a subgraph of G rends to 1 for $p(n) = O\left(n^{\frac{-v}{e}}\right)$. Therefore, $p(n) = n^{\frac{-v}{e}}$ is a threshold, as we wanted to prove. □

Notice that we used exactly the same arguments as we did in K_4, but in this case the count of edges and vertices is slightly more difficult. We can generalize even more this theorem, so that we can compute a threshold not only for K_k. We state the threshold function in case the subgraphs we are looking for are balanced graphs. First we define this concept and then we state and prove the main theorem:

Definition 4.6. Let G be a graph. Recall that $\rho(G) = \frac{e_G}{|G|}$. Then,

- G is *balanced* if for all subgraph $H \subset G$, $\rho(H) \leq \rho(G)$.
- G is *strictly balanced* if for all subgraph $H \subset G, \rho(H) < \rho(G)$.

Notice that K_n is a balanced graph for every n. In particular, it can be clearly seen in figure 1 for $n = 4$. Let us show an example of no balanced subgraph in the following figure 4

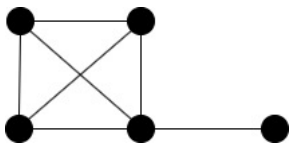


Figure 4: Example of no balanced graph.

Notice that the density of the subgraph K_4 is $\frac{6}{4} = \frac{3}{2}$, and the density of the whole graph is $\frac{7}{5}$, and since $\frac{7}{5} < \frac{3}{2}$, it breaks the condition to be a balanced graph. Once the definition is clear, let us now state the threshold function for containing a balanced subgraph:

Theorem 4.7. *[AS04]* Let H be a balanced graph, $v = |H|, e = e_H$. Then, the property *"To contain H"* for any random graph $G(n,p)$ has threshold function $p = n^{\frac{-v}{e}}$.

Proof. We follow the same structure of the proof for K_k. Let X be the number of copies of H contained in G. If S is a set of v vertices in G, we consider the indicator random variable X_S for S to be a copy of H. Calculating the expectation,

$$\mathbb{E}[X] = \sum_{|S|=v} X_S = \binom{n}{v} p^e \approx n^v p^e.$$

Taking $p(n) = o\left(n^{\frac{-v}{e}}\right)$, $\mathbb{E}[X] \approx n^v n^{-v-\varepsilon} = n^{-\varepsilon} \to 0$, as $\varepsilon > 0$ and $n \to \infty$. We have $X \geq 0$ and $\mathbb{E}[X] = 0$, so X must be 0 *whp*. Therefore, for $p(n) = o\left(n^{\frac{-v}{e}}\right)$ there is not any copy of H contained in G *whp*.

As happened with K_k, in the case of taking $p(n)O\left(n^{\frac{-v}{e}}\right)$, clearly $\mathbb{E}[X] \to \infty$, but notice that it does not imply that $X > 0$. We have seen that the expected value for H is $n^v p^e$, so $\mathbb{E}[X]^2 = n^{2k} p^{2e}$. Let us compute the expected number of intersections between different copies of H into G (considering intersections that share l vertices, for $l = 2, \dots, v-1$).For every l, we fix the l vertices and the r edges joining them (asymptotically $\approx n^i p^r$, with $r = 0, \dots, l-1$) and we choose the rest of vertices and edges to complete the two copies of H. So we get

$$\mathbb{E}[X] = n^l p^r (n^{v-l} p^{e-r})^2 = n^{2v-l} p^{2e-r}.$$

Let us compute the limit to check that it is $o(\mathbb{E}[X]^2)$, if H is balanced:

$$\lim_{n\to\infty} \frac{n^{2v-l} p^{2e-r}}{n^{2v} p^{2e}} = \lim_{n\to\infty} n^l p^r - 1.$$

Hence, since we are now considering $p = O\left(n^{\frac{-v}{e}}\right)$, then $n^l p^r = O\left(n^l n^{-\frac{vr}{e}}\right) = n^{i-\frac{vr}{e}}$. At this point, as H is balanced, by definition $\frac{i}{r} > \frac{v}{e}$. It implies that $i - \frac{vr}{e} > 0$, so $(n^l p^r)^{-1} \to 0$. By definition, it means that $o(\mathbb{E}[X]^2)$. Following as we did for the case $H = K_k$, the sum Δ is also $o(\mathbb{E}[X]^2)$. Because of 4.3, it implies that $X > 0$ *whp*, so the probability of having a copy of H as a subgraph of G tends to 1 for $p(n) = O\left(n^{\frac{-v}{e}}\right)$. Therefore, $p(n) = n^{\frac{-v}{e}}$ is a threshold, as we desired to show. □

The structure in the proof is similar to the threshold function for containing K_k, and notice the importance of H to be balanced. Without this property, it is impossible to get this threshold for any H. Not only K_k is balanced, some other important graphs are also balanced. Hence, their threshold functions can be obtained from theorem 4.7. For instance, we state it for the particular cases of trees and the graph cycle C_n. In particular, notice that two important structures such as trees and cycles are balanced graphs. For trees, its threshold function is the following

Corollary 4.8. The property *"To contain a tree of k vertices"* for any random graph $G(n,p)$ has threshold function $p = n^{\frac{-k}{k-1}}$.

Moreover, for C_k the threshold is independent of the length of the cycle (because it has k vertices and k edges), so cycles of any length show the same asymptotic behaviour with respect to $p(n)$:

Corollary 4.9. The property *"To contain a C_k, for $k \geq 3$"* for any random graph $G(n,p)$ has threshold function $p = \frac{1}{n}$.

Now, all this results allows us to take a general overview of the behaviour of a random graph for different values of $p \in [0,1]$. Obviously, for $p = 0$ the graph consists in n vertices without any edge. The first structures that appear when p rises from 0 are trees. We have seen that trees of k vertices appear with probability 1 when $p = O\left(\frac{k}{k-1}\right)$. So $G(n,p)$ looks as a forest of different trees of k vertices in this asymptotic value of p. Increasing p, we increase k. When p becomes larger than $\frac{1}{n}$, cycles of any length appears in $G(n,p)$ with probability 1, so the shape of the graph is so different than before. The objective now is to characterize the local behaviour of $G(n,p)$ when $p = \frac{C}{n}$, for C constant. Informally speaking, we want to understand how is the transition from the first type of graph (the forest) to the other one (the graph which contains cycles of all length). In order to achieve this objective, we use the following theorem:

Theorem 4.10. [Brun's sieve] Suppose that there is a constant μ such that $\mathbb{E}[X] \to \mu$ and such that for every fixed r, $S^{(r)} = \mathbb{E}\left[\binom{X}{r}\right] \to \frac{\mu^r}{r!}$. Then,

$$\mathbb{P}(X = k) \to \frac{\mu^k}{k!} e^{-\mu}, \text{for all } k \geq 0.$$

Proof. Let us start with the case $k = 0$. Consider $\varepsilon > 0$. Using Taylor expansion, there exists s such that

$$\left| \sum_{r=0}^{2s} (-1)^r \frac{\mu^r}{r!} - e^{-\mu} \right| \leq \frac{\varepsilon}{2}.$$

Because of the inclusion-exclusion principle, clearly $P(X = 0) = 1 - S^{(1)} + S^{(2)} - S^{(3)} + ...$ By the Bonferroni inequalities, since 2s is even, for all value of s we have

$$\mathbb{P}(X = 0) \leq \sum_{r=0}^{2s} (-1)^r S^{(r)}.$$

Moreover, by assumption, there exists n_0 such that for all $n \geq n_0$, and $0 \leq r \leq 2s$ we have

$$|S^{(r)} - \frac{\mu^r}{r!}| < \frac{\varepsilon}{2(2s+1)}.$$

Then,

$$\mathbb{P}(X = 0) = \left(\mathbb{P}(X = 0) - \sum_{r=0}^{2s} (-1)^r S^{(r)} \right) + \left(\sum_{r=0}^{2s} (-1)^r (S^{(r)} - \frac{\mu^r}{r!}) \right) + \left(\sum_{r=0}^{2s} (-1)^r \frac{\mu^r}{r!} - e^{-\mu} \right)$$

$$+e^{-\mu} \leq 0 + (2s+1)\frac{\varepsilon}{2(2s+1)} + \frac{\varepsilon}{2} + e^{-\mu} = e^{-\mu} + \varepsilon.$$

On the other hand, we use the same arguments with the lower bound given by Bonferroni inequality (taking the sum until $2s+1$, which is always odd). We obtain

$$\mathbb{P}(X=0) \geq e^{-\mu} - 0 - 2s\frac{\varepsilon}{2(2s+1)} - \frac{\varepsilon}{2} \geq e^{-\mu} - \varepsilon.$$

Since ε is arbitrary, $\mathbb{P}(X=0) \to e^{-\mu}$.

For the general case, the idea is the same but the expressions are more complicated. As before, we work in detail the upper bound, since the calculus for the lower bound are analogous. We fix $k > 0$, $\varepsilon > 0$. Using Taylor expansion, let us choose s such that

$$\left| \sum_{r=0}^{2s} (-1)^r \binom{k+r}{k} \frac{\mu^{k+r}}{k+r!} - \frac{\mu^k}{k!} e^{-\mu} \right| \leq \frac{\varepsilon}{2}.$$

Using Bonferroni inequalities as before, there exists n_0 such that for all $n \geq n_0$ and $0 \leq r \leq 2s$ we have

$$\binom{k+r}{k} \left| S^{(k+r)} - \frac{\mu^r}{r!} \right| \leq \frac{\varepsilon}{2(2s+1)}.$$

Then, exactly as we did for the case $t = 0$, we put together all this inequalities and we get

$$\mathbb{P}(X=k) = \left(\mathbb{P}(X=k) - \sum_{r=0}^{2s} (-1)^r \binom{k+r}{k} S^{(k+r)} \right)$$

$$+ \left(\sum_{r=0}^{2s} (-1)^r \binom{k+r}{k} \left(S^{(k+r)} - \frac{\mu^{k+r}}{(k+r)!} \right) \right) + \left(\sum_{r=0}^{2s} (-1)^r \binom{k+r}{k} \frac{\mu^{k+r}}{(k+r)!} - \frac{\mu^k}{k!} e^{-\mu} \right)$$

$$+ \frac{\mu^k}{k!} e^{-\mu} \leq 0 + (2s+1)\frac{\varepsilon}{2(2s+1)} + \frac{\varepsilon}{2} + \frac{\mu^k}{k!} e^{-\mu} = \frac{\mu^k}{k!} e^{-\mu} + \varepsilon.$$

For the lower bound, taking as before the Bonferroni inequality for the sum until $2s+1$, we get

$$\mathbb{P}(X=k) \geq \frac{\mu^k}{k!} e^{-\mu} - \varepsilon.$$

Since ε is arbitrary, $\mathbb{P}(X=k) \to \frac{\mu^k}{k!} e^{-\mu}$, as we wanted to show. □

At this point, we prove that the conditions for the Brun's sieve are accomplished for $G(n,p)$ for the threshold function of the property *"To contain a subgraph which is a cycle"*. Recall that it is $p = \frac{C}{n}$, for C constant. Notice that this value of p comes imposing that $\mathbb{E}[X] = o(1)$ in theorem 4.7, so the difficulty comes when proving that $\mathbb{E}\left[\binom{X}{r}\right] \to \frac{C^r}{r!}$ for every fixed r. The following result states that the appearance of cycles when p rises in the asymptotic window of $\frac{1}{n}$ follows a Poisson distribution:

Theorem 4.11. Given a random graph $G(n,p)$, let X be the random variable which counts the number of copies of the graph cycle of length $k \geq 3$ in $G(n,p)$. Then, if $p = \frac{C}{n}$ with C constant, $X \sim Po(C)$.

Proof. Let k be a fixed integer $k \geq 3$, and X the random variable defined in the theorem. In the proof of theorem 4.7 we choose the threshold function for X in a way such that $\mathbb{E}[X] = o(1)$, so $\mathbb{E}[X] \to C$ for a constant value $C > 0$. Let us compute

$$\mathbb{E}\left[\binom{X}{r}\right] = \mathbb{E}\left[\frac{X(X-1)\dots(X-r+1)}{r!}\right] = \mathbb{E}[X^r] + o(\mathbb{E}[X^r]).$$

Let us consider

$$X = \sum_{i \in V, |i|=k} X_i,$$

where every X_i is the indicator random variable for every i set of k vertices to be a cycle. Using this descomposition of X, then

$$\mathbb{E}[X^r] = \sum_{X_i \in V} \mathbb{P}(X_1 \cap \dots \cap X_r).$$

At this point, let us show that in fact $\mathbb{E}[X^r]$ can be deduced only using the set of X_i which are disjoints. Indeed, if all X_i are disjoints

$$\mathbb{P}(X_1 \cap \dots \cap X_r) = \mathbb{P}(X_1)\dots\mathbb{P}(X_r) = \frac{\mathbb{P}(X_1)^r}{r!} = \frac{C^r}{r!}.$$

Similarly, if there is at least an edge in the intersection between X_i, X_j then

$$\mathbb{P}(X_1 \cap \dots \cap X_r) = \frac{(n^s p^s)^r}{r!}.$$

Since $s < 1$, this is $= o(1)$, when $n \to \infty$. Hence, $\mathbb{E}\left[\binom{X}{r}\right] \to \frac{C^r}{r!}$, and by Brun's sieve (Theorem 4.10) we obtain that $\mathbb{P}(X = k) \to \frac{C^k}{k!}$. Since k is arbitrary, X follows a Poisson distribution with parameter C, as we wanted to prove. □

The study done only accounts for balanced subgraphs, but there are also results about the same topic for any subgraph, either balanced or not. The main result in this field, shown in [Bol01], states that when looking for the appearance of any subgraph the most dense balanced subgraph inside it is the key point. Let us consider X_H the number of copies of a subgraph H in $G(n,p)$. Then, the random variable

$$\frac{X_H - \mathbb{E}[X]_H}{\sqrt{\mathbb{V}\mathrm{ar}[X_H]}}$$

tends in distribution to a Normal law if and only if both np^m and $n^2(1-p)$ tend to infinity, where $m = \max\{\rho(H') : H' \subset H\}$. Notice that this result is a generalization for all subgraph,

so in particular it is also true for H balanced. This concludes the study of this graph property. In the following chapters, we study in detail the threshold functions for other two important properties in a graph: connectivity and hamiltonicity. Joining this results with the ones achieved in this chapter, we aim to get a good understanding of random graphs, focusing on its main properties.

5. The birth of the giant component

In this section, we work the threshold function not for the property *"To be connected"* but for the property *"To contain a connected component of linear size in n"*. In spite of this fact, the property we prove is enough to get pretty good results. This property was first studied by Erdős and Rényi in the paper [ER59] . They proved that $G(n,p)$ experiments a sharp phase transition around the threshold function $p(n) = \frac{1}{n}$. The statement of this theorem, usually referred to as *The birth of the giant component*, is the following:

Theorem 5.1. *[ER59]* For any fixed $\varepsilon > 0$, if $p = \frac{1-\varepsilon}{n}$ then $G(n,p)$ has with high probability all connected components of size at most logaritmic in n. If $p = \frac{1+\varepsilon}{n}$, then $G(n,p)$ has a unique component of linear size in n (that is the giant component).

Notice that this result can be not intuitive. When p rises from 0 to 1, clearly more edges are added to the graph, increasing the number of vertices connected. One can think that the small connected components when p is small are merged in connected components of bigger size, so the number of connected components decreases in a smooth way. However, this theorem states the opposite fact. Most of this small connected components end up forming a big connected component, while the rest remains small.

Erdős and Rényi worked with the $G(n,M)$ model in all the paper, so they found this result using counting arguments. However, we use the $G(n,p)$ model, so we work using probabilistic arguments. We show two different proofs for this result, based in completely different approaches. Moreover, the two approaches give different details that complements the main results. The first one is based on *Depth First Search* (*DFS*), a graph search algorithm, and also states the existence of a path of linear length in n inside the giant component. The second one, based in *Galton-Watson branching processes*, does not provide the existence of that path but bounds the size of the second largest component in $G(n,p)$.

5.1 *DFS* approach

In the first proof we show, found in [KS13], we also prove that the giant component in fact contains a path of length linear in n. This statement was also proved by Ajtai, Komlós and Szemerédi in 1981 [AKS81] using a more complex proof. Summarizing, in this section we look at a simpler proof, based in the *DFS* algorithm, of the original results that also gives the Szemerédi and al. result.

Let us introduce the main concepts and notation of the *DFS* algorithm. The *Depth First Search* (*DFS*) is a graph search algorithm that visits all vertices of a graph G. It starts in a root node and explores its neighbors as far as possible before backtracking, and it iterates this process until all vertices have been seen. More formally, when the algorithm is working it divides the vertices of G, (we call them V) on three sets, named S, T and U, defined as follows:

- S is the set of vertices whose exploration is complete, so a vertex in S does not have unexplored neighbors.
- T is the set of vertices not seen yet by the algorithm.
- U is the set of vertices (organized as a stack structure, meaning that the last element that joins the stack is the first that goes out) that the algorithm is exploring at the moment.

There is also an order σ that makes the algorithm choose the vertices following it, but this order does not affect the general behaviour for the things we are looking for.

The *DFS* starts with $S = U = \emptyset$, and $T = V$, and it runs until $U \cup T = \emptyset$. At each round of the algorithm, if $U = \emptyset$ it picks the first vertex in T according to σ and push it into U. So if there is any vertex in U, it takes the last one v (remember the stack structure) and looks in T for neighbors of v. If there is not any neighbor, v is popped out of U and moved to S. On the other hand, if v has some neighbors, the first one (as always according to σ) is pushed from T to U, and the algorithm continues iterating on this new vertex.

Let us remark several observations on this method. First of all, when U is empty and we add a vertex of T, all the vertices that comes in U before it gets empty again are part of the same connected component. The time between two consecutive emptyings of U is called an *epoch*. So each epoch correspond to a connected component of G. Notice also that at each round one vertex is moved. It can be moved either from T to U, or from U to S. Moreover, at each round there is no edges between vertices of S and T by construction of the algorithm. Finally, the set U always spans a path, as a new vertex added to U augment the path spanned until the last vertex before it. Let us show in a simple example how this algorithm works. The graph taken is the following:

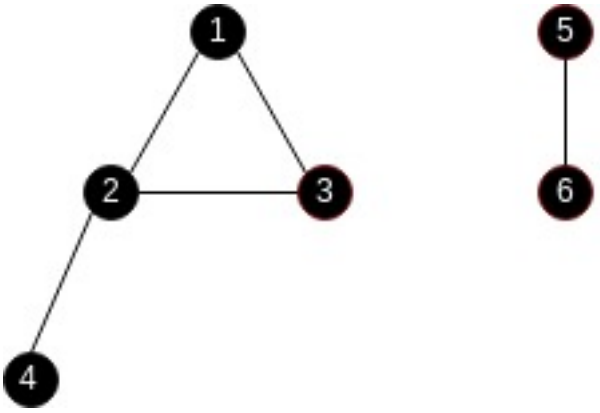


Figure 5: Example graph to apply DFS

Let us explain how the algorithm would explore the graph step by step. Recall that vertices are labelled as in figure 5:

- **The algorithm starts:** We pick a non-visited vertex (vertex 1) and we pass it from T to U.
- **Step 1:** We explore adjacent vertices of vertex 1. We pick vertex 2 and we pass it from T to U.
- **Step 2:** We explore adjacent vertices of vertex 2. We pick vertex 4, and we pass it from T to U.
- **Step 3:** As vertex 4 has no adjacent vertices, we pass vertex 4 from U to S.
- **Step 4:** We pass vertex 3 from T to U, because it is adjacent to vertex 2.
- **Steps 5,6,7:** We pass vertices 3,2,1 (in this order) from U to S, because they do not have unexplored adjacent vertices. End of first epoch, the first connected component has been explored.
- **Step 8:** We pick vertex 5, and we pass it from T to U.
- **Step 9:** We explore adjacent vertices of vertex 5. We pass vertex 6 from T to U
- **Steps 10,11:** We pass vertices 6,5 (in this order) from U to S, because they do not have unexplored adjacent vertices. The exploration is complete

Let us illustrate every step in the following figures, in order to understand completely the performance of the algorithm. We have joint some steps (such as 10,11), as in the explanation, to be more concise:

Figure 6: The algorithm starts

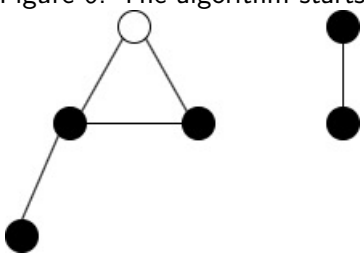


Figure 7: Step 1

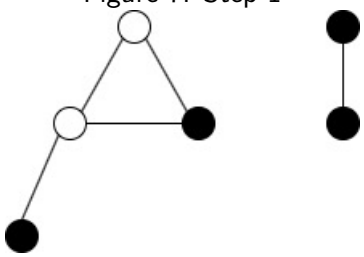


Figure 8: Step2

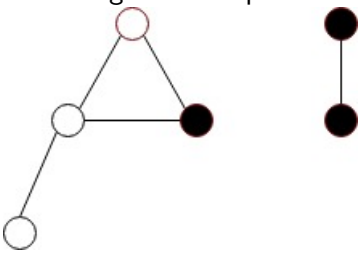


Figure 9: Step 3

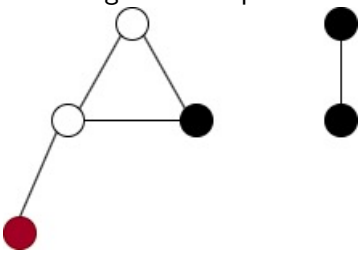


Figure 10: Step4

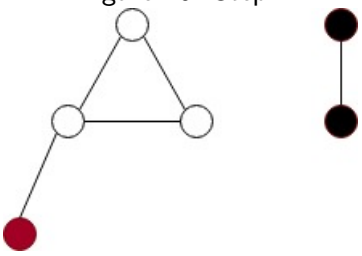


Figure 11: Steps 5,6,7. End of first epoch

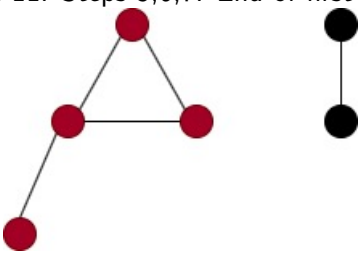


Figure 12: Step 8

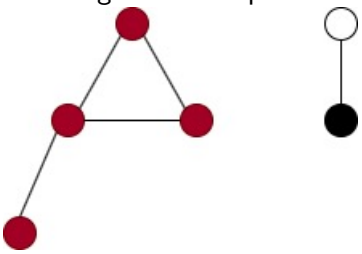


Figure 13: Step 9

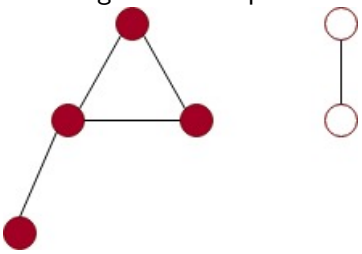


Figure 14: Steps 10,11. The algorithm ends

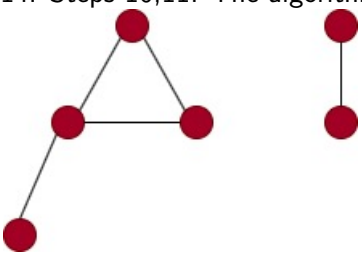


Figure 15: Example of *DFS* applied to a simple graph. For every step, the set S (visited vertices) is painted in red, the set T (unvisited vertices)is painted in black, and the set U (currently visiting vertices) is painted in white.

We have seen the idea and some properties of the *DFS*, but at this point we do not know how to run it on a random graph $G(n, p)$. The idea to perform this process is to run the *DFS* on a sequence of independent and identical random variables $X_i \sim Bern(p), i = 1, \dots, \binom{n}{2}$. Then, at each round of the algorithm, on the one hand if $X_i = 1$ then we consider that the

i-th query of the algorithm is answered as positive. So a vertex is moved from T to U unless $T = \emptyset$. In particular, we observe that after t queries we have $|S \cup U| \geq \sum_{i=1}^{t} X_i$. Furthermore, at time t, $|U| \leq 1 + \sum_{i=1}^{t} X_i$

On the other hand, if $X_i = 0$ it is answered as negative. Clearly, this process leads to a $G(n,p)$ structure. Thus, studying the properties of this sequence X gives all the information of the connected components of $G(n,p)$. In order to prove the main theorem, we need this previous lemma:

Lemma 5.2. *[KS13]* Let $\varepsilon > 0$ constant. Consider the sequence $X = X_i, i = 1 \dots n$ of *i.i.d* Bernoulli random variables X_i, with parameter p. Then:

1. Let $p = \frac{1-\varepsilon}{n}$. Let $k = \frac{7}{\varepsilon^2} \log n$. Then with high probability there is no interval of length kn in [N], in which at least k of the X_i take value 1.
2. Let $p = \frac{1+\varepsilon}{n}$. Let $N_0 = \frac{\varepsilon n^2}{2}$. Then, *whp* $\left| \sum_{i=1}^{N_0} X_i - \frac{\varepsilon(1+\varepsilon)n}{2} \right| \leq n^{\frac{2}{3}}$.

Proof. We prove each point separately.

1. As X_i are Bernoulli random variables with parameter p, the sum for a given interval I of length kn is distributed as a binomial $B(kn,p)$. The probability for a interval of length kn to have at least k random variables $X_i = 1$ can be written as $\mathbb{P}(B(kn,p) \geq k)$. By Chernoff's inequality, we obtain that

$$(N-k+1)\mathbb{P}(B(kn,p) \geq k) < n^2 e^{\frac{\varepsilon^2(1-\varepsilon)k}{3}} < n^2 e^{\frac{\varepsilon^2(1-\varepsilon)7\log n}{3\varepsilon^2}} = o(1),$$

   for $\varepsilon > 0$ small enough. So from this inequality we deduce the desired result.

2. As before, the sum of Bernoulli is distributed as a Binomial random variable. In this case, $\sum_{i=1}^{N_0} X_i \sim B(N_0,p)$. Hence, its expectation is $N_0 p = \frac{\varepsilon n^2 p}{2} \frac{\varepsilon(1+\varepsilon)n}{2}$. Moreover, its standard deviation is $\sigma = \frac{\varepsilon(1+\varepsilon)n}{2}(1-p) \approx cn$, so it is of linear order. Applying Chebyshev inequality we get

$$\left| \mathbb{P}\left(\sum_{i=1}^{N_0} X_i - \frac{\varepsilon(1+\varepsilon)n}{2} \geq (cn)^{\frac{2}{3}} \right) \right| < \frac{1}{cn^{\frac{4}{3}}} = o(1),$$

   so thanks to this inequality we get the desired result. □

Once this lemma is proved, we can state the main result, a variant of Theorem 5.1 which includes also the existence of a path of linear length in n:

Theorem 5.3. *[KS13]* Let $\varepsilon > 0$ be a small enough constant. Let $G \sim G(n,p)$. Then:

1. If $p = \frac{1-\varepsilon}{n}$, then *whp* all connected components of G are of size at most $\frac{7}{\varepsilon^2}\log n$.
2. If $p = \frac{1+\varepsilon}{n}$, then *whp* G contains a path of length at least $\frac{\varepsilon^2 n}{5}$.

Proof. We prove each point separately.

1. We assume that G has a connected component C with more than $k = \frac{7\log n}{\varepsilon^2}$, and we proceed by looking for a contradiction. Let us consider the epoch of the *DFS* where C was created. In particular, we consider the moment when the algorithm has found the $(k+1)$ vertex of C and is going to move it to U. Then, if we denote $\Delta S = S \cap C$, clearly $|\Delta S \cup U| = k$. Hence, the *DFS* has revealed k random variables X_i such that $X_i = 1$, and each positive answer corresponds to a vertex of C (except the first one, which was put when U was empty). The algorithm has looked at edges touching vertices of ΔS or U. Therefore, the number of X_i the algorithm has been queried is less or equal than $\binom{k}{2}$ $+k(n-k)$. This is because $|\Delta S \cup U| = k$, so there are $\binom{k}{2}$ possible edges, and there are $k(n-k)$ possible edges between $\Delta S \cup U$ and the rest of vertices in the graph.
   Since $\binom{k}{2} + k(n-k) < kn$, there is an interval I of length $|I| < kn$, such that $\sum_{i \in I} X_i = k$, so at least k random variables X_i take value 1. This is a contradiction with the property 1 of lemma 5.2
2. Since $p = \frac{1+\varepsilon}{n}$, we assume the property 2) of lemma 5.2 is fulfilled. Let us show that after $N_0 = \frac{\varepsilon n^2}{2}$ queries of the algorithm, U contains at least $\frac{\varepsilon^2 n}{5}$ vertices, so they span a path of linear length in n.
   First, we see that $|S| < \frac{n}{3}$ at time N_0. We proceed again by contradiction. If $|S| \geq \frac{n}{3}$, we consider the moment t where $|S| = \frac{n}{3}$. At that moment, $|U| \leq 1 + \sum_{i=1}^{t} X_i$, by construction of the algorithm. Because of property 2) of Lemma 5.2,

$$1 + \sum_{i=1}^{t} X_i < \frac{n}{3}.$$

   Then, $|T| \geq \frac{n}{3}$, since $|T| + |S||U| = n$. So the algorithm has looked for all possible edges between S and T. There are $|S||T| \geq \frac{n^2}{9} > N_0$ possible edges, and the algorithm has found them to be non-edges, by definition of S and T. This is clearly a contradiction, as the number of queries can not be greater than N_0. So we assume that in time N_0, $|S| < \frac{n}{3}$. Again by contradiction, we prove that $|U| \geq \frac{\varepsilon^2 n}{5}$. If $|U| < \frac{\varepsilon^2 n}{5}$ and $|S| < \frac{n}{3}$, then $|T| \neq \emptyset$. So a positive answer to a query of the *DFS* results on moving a vertex

from T to U.

By property 2) of Lemma 5.2 the value of $\sum_{i=1}^{N_0} X_i$ is at least $\frac{\varepsilon(1+\varepsilon)n}{2} - n^{\frac{2}{3}}$.

Since $\sum_{i=1}^{N_0} X_i$ is the number of positive answers to the algorithm, by construction $|S \cup U| \geq \frac{\varepsilon(1+\varepsilon)n}{2} - n^{\frac{2}{3}}$. Since $|S \cup U| - |U| + |S \cap U| = |S|$ and $|U| < \frac{\varepsilon^2 n}{5}$, then

$$|S| \geq \frac{\varepsilon n}{2} + \frac{\varepsilon^2 n}{2} - n^{\frac{2}{3}} - \frac{\varepsilon^2 n}{5} + 0 = \frac{\varepsilon n}{2} + \frac{3\varepsilon^2 n}{10} - n^{\frac{2}{3}}.$$

All queries asking for pairs between S and T have been answered as negative by the algorithm. They are $|S||T| \geq |S|(n - |S| - \frac{\varepsilon^2 n}{5})$ queries. Hence, we get

$$\frac{\varepsilon n^2}{2} = N_0 \geq |S||T| \geq |S| \left(n - |S| - \frac{\varepsilon^2 n}{5} \right) \geq \left(\frac{\varepsilon n}{2} + \frac{3\varepsilon^2 n}{10} - n^{\frac{2}{3}} \right) \left(n - \frac{\varepsilon n}{2} - \frac{\varepsilon^2 n}{2} + n^{\frac{2}{3}} \right).$$

We develop this expression, obtaining:

$$\frac{\varepsilon n^2}{2} + \frac{\varepsilon^2 n^2}{20} - O(\varepsilon^3) n^2 > \frac{\varepsilon n^2}{2} = N_0.$$

The contradiction is clear, as $N_0 > N_0$ can not be possible, so in fact $|U| \geq \frac{\varepsilon^2 n}{5}$, and it concludes the proof. □

At this point, let us comment briefly about connectivity of random graphs. Since it is a monotone property, a threshold function must exist. There is a well-known result, whose proof can be found in [Bol01], that states the value of this threshold function as $\frac{\log n}{n}$. Notice that this threshold function is asimptotically greater than the threshold for the birth of the giant component, so it does not contradict the main result of this section. In the following chapter, let us state and proof another version of theorem 5.1 based on *Galton-Watson* branching processes

5.2 Galton-Watson approach

In this section, we explore a completely different method to find the threshold function for the giant component. The main idea is to iterate a *Galton-Watson* process to look for a spanning tree composed by vertices of the giant component. Notice that in this case we are not proving that the giant component contains a path with length linear in n, as we did in the previous proof. However, in this formulation of the theorem we also give an upper bound for the size of the second biggest component, which is a novelty with respect the previous approach. What's more, the clever idea and the utility of the concepts shown in this demonstration makes this section worth enough to be included in this work.

Let us first present the crucial concept in this section: *Galton-Watson branching process*. It was first introduced by Francis Galton at the end of the XIX century, when he was investigating how to model the extinction of English aristocratic family names. In the simplified Galton-Watson model, we assume that a population evolves in *generations*, and the number of members in the n-th generation is equal to a random variable Z_n. Each individual in the n-th generation gives rise to some members (called *children*) of the $(n+1)$-th generation following a random variable X. An important fact here is that the rise of the families of two different members of the same generation is independent, so the offspring for everybody in this process follows the same distribution X.

If we look at this stochastic process sequentially, we consider a single node v that produces children according to the random distribution X. Each v_i of the set of children produces more children according to a random variable X_i, which follows the same distribution X, independently from the number of children of his parent v, and from the another members of his same generation. Then, we iterate this process, and the question that arises naturally (indeed is the main question why this process was created and studied, remember that Galton's objective was to study the extinction of family names) is if at some point the process finishes, meaning that all random variables of a generation give 0, so the family name become extinct.

More formally, we are considering an infinite sequence of random variables $X_1, \dots, X_i, \dots$, following the same probability distribution X and all of them independent. As we are looking at the vertices of a random graph $G(n,p)$, and they are labelled, we order all vertices in a generation by label, and the i-th member has only one shot of the random variable X_i. Its result is its number of children. As notation, we say that a member of a generation that has already taken the shot of the corresponding random variable is *dead*, and if he has not taken the shot yet we say it is *alive*.

We denote the extinction probability of the process (created following a distribution X) as

$$\rho_X = \mathbb{P}(Y_t = 0), \textit{ for any } t > 0.$$

Let us now study the behaviour of ρ_X. The first vertex of the process produces i children with

probability $\mathbb{P}(X = i)$, and each child has a probability ρ_X (same for all) to end the process. As all X_i are independent, the probability of the process to die out conditioned on $X_0 = i$ is the product of all probabilities, that is ρ_X^i. We sum up for all i and we get the following expression for the probability of extinction starting on the first vertex of the process:

$$\rho_X = \sum_{i=0}^{\infty} \mathbb{P}(X = i)\rho_X^i$$

This expression of the right is indeed the probability generating function of X, which we denote as $f_X(x)$. Without entering in detail, it is a power series representation of the probability mass function of a random variable. It is a very powerful tool, since it allows to apply results of power series theory to random variables. Indeed, the main theorem which gives the extinction probability for any process is proven using the properties of $f_X(x)$. This theorem states that if the expectation of the random variable X is less or equal than 1, the sequence will be extinct *whp*. On the other hand, if this expectation is greater or equal than 1, the theorem gives a formula to compute that probability. The statement of the theorem is the following:

Theorem 5.4. In the previous context, the following results hold:

1. $\mathbb{E}[X] \leq 1$ implies that $\rho_X = 1$ unless $\mathbb{P}(X = 1) = 1$
2. $\mathbb{E}[X] > 1$ implies that ρ_X is the unique solution of the equation $x = f_X(x), x \in [0, 1)$

Let us outline, without entering in detail, the particular cases of extinction probability when $X \sim Po(C)$, and $X \sim Bin(n, p)$, because it is necessary to prove the main theorem in this section. In the first case, when $X \sim Po(C)$, $\rho_X = e^{-\beta(C)C}$, where $\beta(C) = 1 - \rho_X$ is the survival probability. From here, doing some basic calculus it can be seen that the Poisson process ends *whp* if $C < 1$, and if $C > 1$ then ρ_X is the unique solution of the function $f(y) = 1 - y - e^{-cy}$. In the second case, when $X \sim Bin(n, p)$, the extinction probability ρ_X converges to $1 - \beta(C)$, the extinction probability defined by $Po(C)$. This is a key fact in the forthcoming proof.

Let us now outline the proof of the main theorem. In order to study the size of the biggest component in a random graph $G(n, p)$, we pick a vertex v_0 and we generate a spanning tree in its component using a kind of *Galton-Watson* stochastic process. First, we let v_0 randomly select his neighbors $v_1, \dots, v_{X_0}$ (or children, following the *Galton-Watson* point of view) according to a random variable $X_0 \sim Bin(n-1, p)$. Hence, we take v_1 and we randomly select its neighbors from the set of all vertices that are not yet involved in the process, so it is a random variable $X_1 \sim Bin(n - X_0 - 1, p)$. Then, following the same reasoning for v_2, we select its neighbors following $X_2 \sim Bin(n - X_0 - X_1 - 1, p)$. Iterating this process through many generations, we select for the i-th vertex its neighbors within all the still isolated ones, as many as a realization of the random variable $X_i \sim Bin(n - \sum_{j=0}^{i} X_j - 1, p)$. Notice that this process is not the *Galton-Watson* process described before, since the distributions are not the same for every vertex,

they depend on the history of the process. But nevertheless, it is not a big trouble, since the difference between $Bin(n,p)$ and $Bin(n-\sum_{j=0}^{i} X_j - 1, p)$ is not important because $\sum_{j=0}^{i} X_j$ is smaller in order than n. Let us state and prove a new version for the theorem of the giant component:

Theorem 5.5. Let $p = \frac{c}{n}$, where $c > 0$ is a constant. Then,

1. if $c < 1$, *whp* the largest connected component of $G(n,p)$ has at most $\frac{4}{(1-c)^2}\log n$ vertices.
2. if $c > 1$, *whp* $G(n,p)$ contains a single giant component of $(1+o(1))\beta n$ vertices. Furthermore, the second largest component has at most $\frac{16c}{(c-1)^2}\log n$ vertices.

Proof. We prove each point separately.

1. Consider $pn = c < 1$ fixed, and let us denote as $C(v)$ the component of vertex v. Let $k = \frac{4}{(1-c)^2}\log n$ be the threshold function we want to prove for the size of the giant component. We fix a vertex v and we create a spanning tree of its component with a branching process. If $C(v)$ contains more than k vertices, it means that at least k children have been created during the realization of the random variables $X_1, \dots X_k$, which correspond to the first k members of the population. Then, taking $X_i^+ \sim Bin(n,p)$, we obtain
$$\mathbb{P}(|C(v)| > k) \leq \mathbb{P}\left(\sum_{i=1}^{k} X_i \geq k\right) \leq \mathbb{P}\left(\sum_{i=1}^{k} X_i^+ \geq k\right).$$
The second inequality comes from the fact that $X_i \sim Bin(n-m_i, p)$, for some $m_i \geq 1$, so taking a $Bin(n,p)$ we are increasing the probability of getting a bigger sum, since we are taking more shots of a $Bern(p)$, so the probability only can grow up. Because of the properties of the Binomial distribution,
$$\sum_{i=1}^{k} X_i^+ \sim Bin(kn, p),$$
with mean $knp = ck < k$ and standard deviation $(1-c)k$. We assume $c > \frac{3}{4}$, but this is not a problem since we are proving a monotone property, so it is also true for $c < \frac{3}{4}$. Now, applying Chernoff inequality:
$$\mathbb{P}\left(\sum_{i=1}^{k} X_i^+ \geq k\right) \leq e^{-\frac{(1-c)^2k^2}{2ck} + \frac{(1-c)^3}{2(ck)^2}} = e^{-k\left(\frac{(1-c)^2(2c-1)}{2c^2}\right)+O(1)} < e^{-\frac{(1-c)^2}{3c}k+O(1)}.$$
This last expression is $O(n^{-\frac{4}{3}})$, because of the value of k. Then, the probability of having at least k children is upper bounded by a family of functions tending to zero, so this probability (recall that it is also the probability to have a connected component with more than k vertices) is zero *whp*.

2. We assume $np = c > 1$, and we denote $k_- = \frac{16c}{(c-1)^2} \log n$ and $k_+ = n^{\frac{2}{3}}$. First, we show that for every vertex v and every k such that $k_- < k < k_+$ there are two options. The first one is that the process started at v terminated within k_- steps, so that $|C(v)| \leq k_-$. The second one is that after the k-th round there are at least $\frac{c-1}{2}k$ alive vertices, so we can continue iterating the process inside $C(v)$. Consider fixed v and k, let $F_{v,k}$ be the event that the process started at v did not terminate within k steps and after the k-th round there are less than $\frac{c-1}{2}k$ alive vertices. If this happens then the first k random variables X_i produced less than $k + \frac{c-1}{2}k = \frac{c+1}{2}k$ children. In particular, recall that $X_i \sim Bin(n - m_i, p)$, with $m_i \leq \frac{c+1}{2}k$. Let us denote $X_i^- \sim Bin(n - \frac{c+1}{2}k)$ Hence, using in the last inequality the same reasoning about the binomial distribution introduced in the first part of the proof for X_i^+ (that is we are adding more trials of a Bernoulli, so the probability only can grow up), we obtain

$$\mathbb{P}(F_{v,k}) \leq \mathbb{P}\left(\sum_{i=1}^{k} X_i \leq \frac{c+1}{2}k\right) \leq \mathbb{P}\left(\sum_{i=1}^{k} X_i^- \leq \frac{c+1}{2}k\right).$$

Since

$$\sum_{i=1}^{k} X_i^- \sim Bin\left(k(n - \frac{c+1}{2}k), p\right),$$

with expectation approximately equal to $knp = ck > k$ and standard deviation approximately equal to $(c - \frac{c-1}{2})k = \frac{c-1}{2}k$, we can apply Chernoff again:

$$\mathbb{P}\left(\sum_{i=1}^{k} X_i \leq \frac{c+1}{2}k\right) < e^{-\frac{(c-1)^2k^2}{8ck}(1+o(1))} < e^{-\frac{(c-1)^2}{9c}k} = n^{-\frac{16}{9}}.$$

This is true for large n. Now, applying the union bound:

$$\mathbb{P}\left(\bigcup_{v \in V} \bigcup_{k=k_-}^{k^+} F_{v,k}\right) < n \cdot n^{\frac{2}{3}} \cdot n^{-\frac{16}{9}} \to 0.$$

At this point, we have proven that for every vertex v either its component $C(v)$ has at most k_- vertices (in which case we call v a small vertex), or its component $C(v)$ has at least k_+ vertices and has at least $\frac{c-1}{2}k$ alive vertices after the k-th round (in which case we call v a large vertex).
The second big step in the proof is to demonstrate that there is a single component that contains all large vertices. Let v', v'' be two large vertices. Let us grow the process from v'. If we reach v'' we are done. Otherwise, we stop after k_+ rounds and we consider a set $L_{v'}$ of live vertices in $C(v')$, with $|L_{v'}| = \frac{c-1}{2}k_+$. Then, we grow the process from v'', and we follow the same reasoning. If we reach v', we are done. Otherwise, we stop after k_+ rounds and we consider a set $L_{v''}$ of live vertices in $C(v'')$, with $|L_{v''}| = \frac{c-1}{2}k_+$. The possible edges between $L_{v'}$ and $L_{v''}$ were not yet explored and if there is any one of those

edges in the graph, then $C(v') = C(v'')$, as we want. The probability to have not edges between $L_{v'}$ and $L_{v''}$ can be computed easily by counting them, since the probability to have not an edge is $(1-p)$ and we know there are $\frac{(c-1)^2}{4}k_+^2$ possible edges. So we obtain the following expression, and we can bound it using the value of k_+:

$$(1-p)^{\frac{(c-1)^2}{4}k_+^2} < e^{-\frac{c(c-1)^2}{4n}n^{\frac{4}{3}}} = e^{-Cn^{\frac{1}{3}}}.$$

So by the union bound any two large vertices are in the same component *whp*, since $1 - n^2 e^{-Cn^{\frac{1}{3}}} \to 1$.

Finally, the last part in the proof consist on to bound the size of the component containing all large vertices. First we estimate the probability of a vertex to be small. Let $\rho(n)$ be the probability of extinction for the *Galton-Watson* process generated with the binomial distribution $Bin(n,p)$. For every vertex v, let us denote by $\nu(n,p)$ its probability of being small. Let us find a lower bound for this probability. By definition, our *Galton-Watson* process makes less trials at every step than a process iterated over a $Bin(n,p)$. So the lower bound for $\nu(n,p)$ is the probability of extinction minus the probability that it dies after k_- steps (we call this event F_{k_-}). Then, we bound this probability:

$$\mathbb{P}(F_{k_-}) \leq \sum_{k \geq k_-} \mathbb{P}(X_1 + \cdots + X_k \leq k) \leq \sum_{k \geq k_-} e^{\frac{(c-1)^2}{c}k} = o(1).$$

In the last inequality we have used Chernoff bound again, and the sum for all this values clearly tends to 0 even summing all terms. On the other hand, we can upper bound it using that if v is small the process starting from v iterates some random variables $Bin(n-m_i,p)$, with $m_i \leq k_-$. So we obtain:

$$\rho(n) - o(1) \leq \nu(n,p) \leq \rho(n-k_-).$$

When n tends to infinity, both $\rho(n)$ and $\rho(n-k_-)$ tend to the same extinction probability $\rho_p = 1-\beta(c)$ of the Poisson process of mean c. Then, the expectation for a single vertex to be small is ρ_p, so by linearity of expectation if we call Y the random variable which counts the number of small vertices in the graph we obtain

$$\mathbb{E}[Y] = (1 - \beta(c) + o(1))n.$$

We now have to prove that Y is concentrated about its mean, in order to ensure that the size of the largest component is *whp* $\beta(c)n$. Let us then calculate the variance of Y. We call S_v the indicator random variable for the event that v is small. Assuming that v is small, let us bound the probability that another vertex u is also small. There are two possible options for this case. The first one is that both are in the same component, and the second one (and most probable, because by definition two small vertices tend to be not connected) is that both are in separated components. Summing the probability of this two options, we get:

$$\mathbb{P}(S_v = 1 | S_u = 1) \leq \frac{k_-}{n} + \nu(n-k_-,p).$$

The first term comes computing the probability for u to be in $C(v)$. For the second term, we get the upper bound growing a spanning tree for the component of u in all $V - C(v)$ vertices and using that by definition $\nu(V - C(V), p) \geq \nu(V, p)$, since if there are more vertices there is less probability for a single one to be small.
Hence, we can compute the variance for Y:

$$\mathbb{V}\mathrm{ar}[Y] = \sum_v \sum_u (\mathbb{E}[S_v S_u] - \mathbb{E}[S_v]\mathbb{E}[S_u])$$

$$= \mathbb{E}[Y] + \sum_v \mathbb{P}(S_v = 1) \sum_{u \neq v} \mathbb{P}(S_u = 1 | S_v = 1) - \mathbb{E}[Y]^2$$

$$\leq \mathbb{E}[Y] + \sum_v \mathbb{P}(S_v = 1)(n-1)\left(\frac{k_-}{n} + \nu(n - k_-, p)\right) - \mathbb{E}[Y]^2$$

$$\leq \mathbb{E}[Y] + \mathbb{E}[Y](k_- + n\nu(n - k_-, p)) - \mathbb{E}[Y]^2 = o(\mathbb{E}[Y]^2)$$

In the last inequality, we have used that $\nu(n - k_-, p) \to 1 - \beta(c)$ and $k_- = o(n)$, so $k_- + n\nu(n - k_-, p) = (1 + o(1))\mathbb{E}[Y]$. Thus, Applying Chebyshev inequality, we get the result we want, since it gives that $Y = (1 - \beta(c) + o(1))n + o(1)$. Then, the expectation for the number of large vertices is $n - \mathbb{E}[Y] = n\beta(c) + o(1)$, and we have proven that all large vertices lie in the same component, so the we can claim that the size of this single giant component is $(1 + o(1))\beta n$, as we wanted to prove. □

This abrupt appearance of the giant component is not an exclusive property of random graphs. It appears in some other discrete structures. Let us show a recently achieved result about the giant component that generalizes the main theorem we have shown. First, we introduce the concept of hypergraph. An hypergraph is a generalization of a graph, where the so-called *hyperedges* can join an arbitrary number of vertices, not only two vertices as edges do in graphs. Notice that the definition of connectivity has to be different respect to the definition on standard graph. In this case, the *k-tuple-connectivity* relation can be defined (for any fixed k) by letting two sets of k vertices be connected if they lie in a common edge and considering the transitive closure. Notice that for $k = 1$ we recover the definition of vertex connectivity for a graph. The most worked hypergraphs are the so-called *k-uniform* hypergraphs, where every hyperedge joins k vertices of the hypergraph. We can see in figure 16 an example of how a *3-uniform hypergraph* looks like.

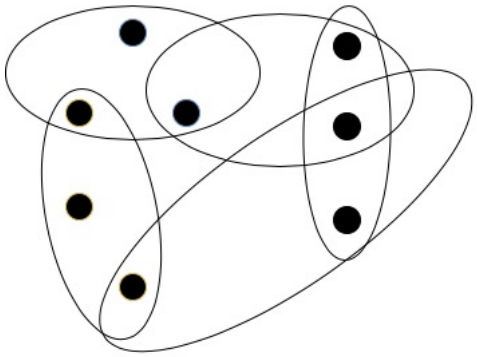


Figure 16: 3-uniform hypergraph with eight vertices and five hyperedges.

There is a wide theory on random hypergraphs. In particular, it is possible to generalize the $G(n,p)$ model to a *k-uniform* hypergraph , considering that every hyperedge has a probability p to exist. In this case, there is also a threshold function for the birth of the giant component. A recent result from 2015, shown in [CKK18] states that the existence of a *j-tuple-connected component* of size linear in the number of *j-sets* shows an abrupt phase transition, with threshold function

$$p = \frac{(k-j)!}{\binom{k}{j} - 1} n^{j-k}.$$

Notice that for $k = 2, j = 1$, the result is the threshold function for the random graph model that we have seen in this work. In the following section we finish the study of threshold functions for graph properties, setting the threshold function for hamiltonicity.

6. Hamiltonicity of $G(n, p)$

In this section, we study the threshold function for the property *"To be hamiltonian"* for a given random graph $G(n, p)$, following the arguments found in [Kri16]. Recall that a graph is hamiltonian if it contains a hamiltonian cycle, i.e. a cycle which passes only once (except for the initial-end vertex, in this case it passes two times, one to start and one to end) for every vertex in the graph. Let us show an example of hamiltonian graph in figure 17:

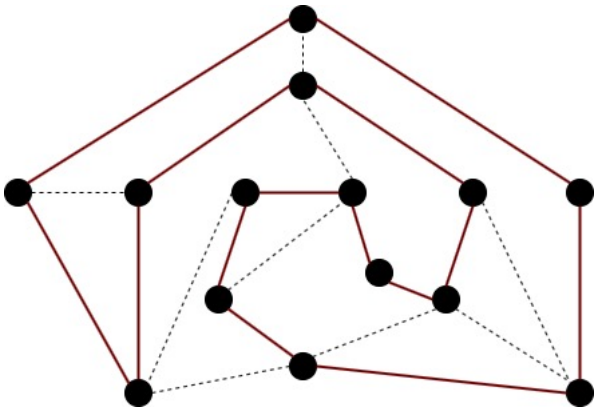


Figure 17: Hamiltonian graph. The hamiltonian cycle is marked in red, the rest of edges are dashed lines

As happened in section 4 with the *clique problem*, to find a hamiltonian cycle in any graph is extremely difficult. Also as in section 4, random graph models allows us to find a threshold function for hamiltonicity, characterizing this property in a much stronger way than for any general graph. In this last case, there are many theorems that impose sufficient conditions for a graph to be hamiltonian. For example, Dirac's theorem [Dir52] states that a simple graph with n vertices is hamiltonian if every vertex has degree $\frac{n}{2}$ or greater. Similarly, Ore's theorem [Ore60] states that a simple graph with n vertices is hamiltonian if for every pair of non-adjacent vertices, the sum of their degrees is greater or equal than n. Notice that this two theorems are based on the degrees of vertices, as the main proof we show in this next section. There are some more theorems similar to the two shown, but for random graphs the study is based again in threshold functions.

Looking at the concept of hamiltonicity, we demand a connected graph, so its threshold function should be asimptotically bigger than the threshold for the connectivity we have seen in the main theorem 5.1 of the previous section, that is $p = \frac{\log n}{n}$. In the proof, we use the fact that a hamiltonian graph does not have vertices of degree one, as it is impossible to get a hamiltonian cycle when a vertex v has only one edge, as we visit its unique adjacent vertex u at least two times if we want to visit v. We claim (and this is the key result) that, in fact, when $\delta(G) \geq 2$ then $G(n, p)$ is *whp* hamiltonian.

6.1 Main result

First, let us set the threshold function for the property $\delta(G) \geq 2$, which coincides at the end with the threshold for hamiltonicity. We claim the result for both $G(n, M)$ and $G(n, p)$, since we use the threshold for $G(n, M)$ to prove the main theorem of this section.

Theorem 6.1. *[Kri16]* Let $w(n)$ be a function such that tends to infinity arbitrarily slowly with n. Then,

1. In the probability space $G(n, p)$,
   (a) if $p = \frac{\log n + \log\log n - w(n)}{n}$, we get $\delta(G) \leq 1$.
   (b) if $p = \frac{\log n + \log\log n + w(n)}{n}$, we get $\delta(G) \geq 2$.
2. In the probability space $G(n, M)$,
   (a) if $M = \frac{n(\log n + \log\log n - w(n))}{n}$, we get $\delta(G) \leq 1$.
   (b) if $M = \frac{n(\log n + \log\log n + w(n))}{n}$, we get $\delta(G) \geq 2$.

Proof. We outline the main ideas in the proof, since it is a standard application of the first and second moment method. Let us consider $X_{\geq 2}$ the random variable which counts the number of vertices of degree greater or equal than 2. We can decompose this random variable as

$$X_{\geq 2} = \sum_{i=1}^{n} V_i,$$

where every V_i is the indicator random variable for the vertex i to have degree greater or equal than two. Of course, all V_i are i.i.d random variables. By linearity of expectation,

$$\mathbb{E}[X_{\geq 2}] = \mathbb{E}\left[\sum_{i=1}^{n} V_i\right] = \sum_{i=1}^{n} \mathbb{E}[V_i] = n\mathbb{E}[V_1].$$

Thus, computing the expected value for a single vertex to have degree 2 or higher, we obtain

$$n\mathbb{E}[V_1] = n\mathbb{P}(\text{vertex 1 has degree} \geq 2) = n(1 - \mathbb{P}(\text{vertex 1 has degree} < 2),$$

which is equal to

$$n(1 - \mathbb{P}(\text{vertex 1 has degree1}) - \mathbb{P}(\text{vertex 1 has degree0})).$$

Writing in terms of p, we get

$$n\left(1 - (1-p)^n - (n-1)p(1-p)^{n-1}\right).$$

Then, using this last expression, it is a standard numerical manipulation to check that $\mathbb{E}[X_{\geq 2}]$ tends to infinity for values of p greater than the one given in the statement of the theorem.

On the other hand, it is analogous to check that $\mathbb{E}[X_{\geq 2}]$ goes to zero for values of p smaller than the one given in the statement of the theorem. This computations conclude the proof, since it is necessary to prove that indeed

$$p = \frac{\log n + \log\log n}{n}$$

is a threshold, as we wanted to prove. □

As said before the statement of this theorem, our goal is to prove that the threshold function for the hamiltonicity is in fact the same as the threshold for $\delta(G) \geq 2$. To achieve this objective, we use the concepts of random graph processes and hitting time for random graphs. Random graph processes are very useful to understand the behaviour of both $G(n,p)$ and $G(n,M)$ from another point of view. $G(n,p)$ comes from a probabilistic view and $G(n,M)$ from a combinatorial one, but a random graph process presents a random graph as an evolution of several steps that we can interpret as timestamps. Let us introduce the main definitions we need to state and prove the main theorem.

Consider a random graph $G(n,p)$ and a random permutation $\sigma : E(K_n) \to [N]$ of all $N = \binom{n}{2}$ possible edges of the random graph (that is the set of edges of the complete graph K_n). Suppose that we take one of all possible random permutations $\sigma = (e_1, \dots, e_N)$ with the same probability. A random graph process $G(\sigma)$ is a sequence of graphs (with n vertices) $G_0, \dots, G_N$, in which the set of edges of G_k, $k = 0, \dots, N$, is exactly the set of the first k edges of the random permutation σ. The sequence $G(\sigma)$ starts with an empty graph and finishes with the complete graph K_n, and by definition we can look at every snapshot G_k as a random graph $G(n,M)$ for $M = k$.

As we said before, we can interpret also a random graph process as an evolutionary process. We start taking the empty graph of n vertices and we take N steps. To generate G_i, we take the $i-th$ edge of the random permutation $\sigma = (e_1, \dots, e_N)$ and we add it to the graph G_{i-1} we got in the previous step. This point of view is very useful to understand the concept of hitting time for a graph property, which is a key element in this work.

Let P be a monotone increasing property of random graphs of n vertices, such that $K_n \in P$. We consider the random graph process $G(\sigma)$ given by a random permutation σ. Then, the hitting time for the property P is the lowest k such that $G_k \in P$. We denote it by $\tau_P(G(\sigma))$. Notice that because of the monotonicity of P, $G_i \notin P$ for all $i = 0, \dots, \tau_P(G(\sigma)) - 1$, and $G_i \in P$ for all $i = \tau_P(G(\sigma)), \dots, N$. Furthermore, notice that the existence of the hitting time is guaranteed by the conditions P increasing and $K_n \in P$. By definition of the random graph process, $\tau_P(G(\sigma)$ is a random variable. Studying $\tau_P(G(\sigma))$ is a very common approach to study the behaviour of a given monotone property P in random graphs. Also it is very common to compare the hitting time for two or more properties, and that is what we do. We claim that the hitting times for the properties $\delta(G) \geq 2$ and G *"To be hamiltonian"* coincide, and from

this fact we deduce that the two properties have the same threshold function. The proof for the following theorem is the most complicated part of this section, so we develop its main ideas later.

Theorem 6.2. *[Kri16]* Let $G(\sigma)$ be a random graph process on n vertices. We denote by $\tau_2(G(\sigma))$ the hitting time for the property $\delta(G) \geq 2$ and $\tau_H(G(\sigma))$ the hitting time for the Hamiltonicity. Then, *whp*

$$\tau_2(G(\sigma)) = \tau_H(G(\sigma)).$$

The main theorem in this section is indeed a corollary of theorems 6.1 and 6.2. In this case, instead of working with $G(n,p)$ as we did in all this work, it is easier to work with $G(n,M)$ and then deduce the result for $G(n,p)$:

Theorem 6.3. *[Kri16]* The following results hold:

1. The property *"To be hamiltonian"* for any random graph $G(n,M)$ has threshold function
$$M = \frac{n(\log n + \log\log n)}{n}.$$
2. The property *"To be hamiltonian"* for any random graph $G(n,p)$ has threshold function
$$p = \frac{\log n + \log\log n}{n}.$$

Proof. We prove each point separately.

1. Let us denote by M the threshold function given in the statement of the theorem. As we discussed before, the $M-th$ step of a random graph process $G(\sigma)$ gives a random graph $G(n,M)$. By theorem 6.1, clearly $\tau_2(G(\sigma)) \leq M$, and by theorem 6.2, $\tau_2(G(\sigma)) = \tau_H(G(\sigma))$, so we deduce $\tau_H(G(\sigma)) \leq M$. That is, the graph process has become hamiltonian *whp* at most at M.
2. We denote by H the property to be hamiltonian. Notice that if we consider $G(n,p)$ and we impose it to have exactly M edges, we get a $G(n,M)$ distribution. Let us take $w_1(n) = \frac{w(n)}{3}$, and consider $I = [Np - nw_1(n), Np + nw_1(n)]$. Because of the threshold for $G(n,M)$, for every $M \in I$, the random graph $G(n,M)$ is *whp* hamiltonian.The number of edges of $G(n,p)$ is a random variable $Bin(n,p)$, so its variance is smaller than Np. Applying Chebyshev inequality we deduce that *whp* the number of edges of $G \sim G(n,p)$ is in the interval I. Then, because of the choose of I,
$$\mathbb{P}(G \notin H) = \sum_{m=0}^{N} \mathbb{P}(|E(G)| = m)\mathbb{P}(G \notin H||E(G)| = m) \leq \mathbb{P}(|E(G)| \notin I)$$
$$+ \sum_{m \in I} \mathbb{P}(|E(G)| = m)\mathbb{P}(G \notin H||E(G)| = m)$$

$$= o(1) + \sum_{m \in I} \mathbb{P}(|E(G)| = m)\mathbb{P}(G \sim G(n,m) \notin H) = o(1),$$

as we desired to prove.

□

At this point, we have supposed as known theorem 6.2 and we have been able to deduce the main result (the threshold function for hamiltonicity). So in order to be coherent we must prove theorem 6.2 to complete all the reasoning.

6.2 Proof of Theorem 6.2. Use of expanders.

First of all, recall that our objective is to prove that $\tau_2(G(\sigma)) = \tau_H(G(\sigma))$. To reach this objective, we have to define some new concepts. Let us start by the most important concept, this is *expanders*.

Expanders are so useful in both mathematics and computer science for many practical applications. Before setting its formal definition, let us say that an expander is a sparse but highly connected graph. Because of its special properties, they are basic when modeling networks, specially biologic and communication ones. They are also important in other mathematical fields, such as code theory or Markov chains. The study of expanders goes to many different directions. For example, the parameters of an expansion and its relationship with other graph invariants, how to efficiently generate a concrete expander, how to test if a graph contains expanders, and more general questions with the final aim of find applications in both theoretical and practical problems, understanding the relations between expansions and other mathematical notions. A good reference in order to dig deeper in this field is [HLW06]. The definition of expander is the following:

Definition 6.4. Given a positive integer k and a positive real α, a graph G is a $(k, \alpha)-$*expander* if $|N_G(U)| \geq \alpha|U|$ for every subset $U \subset V$ of at most k vertices.

To clarify this definition, let us show an example of the $(2,2)-$*expander* condition for a subset U in a graph. Given a graph G, to be a $(2,2)-$*expander* all the subsets of one or two vertices must verify the condition of figure 18. Notice that in this case $|N_G(U)| = 2|U|$. Indeed, it is also accepted that $|N_G(U)| > 2|U|$, this case corresponds to add more vertices in the set $N_G(U)$ in the same figure 18.

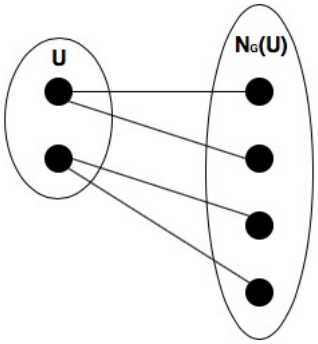


Figure 18: Example of condition to be fulfilled for all set of at most 2 vertices to be a $(2,2)-$ *expander*

The following concept of *booster* is also basic in the proof (indeed boosters allows us to reach hamiltonicity in the forthcoming construction of the proof) but it has not the mathematical importance that expanders have. Let us state the definition of booster:

Definition 6.5. Given a graph G, a non-edge $e = (u, v)$ of G is called a *booster* if adding e to G creates a graph G', which is Hamiltonian or whose longest path is longer than that of G

Finally, we present the set $\mathbb{S}(G)$, which is defined exclusively for the proof. It is build as the set of vertices that have lower degree than a given value, defined in such a way that is useful later in the proof:

Definition 6.6. Given a graph G with n vertices, $\delta_0 > 0$ small enough and $d_0 = [\delta_0 \log n]$, we define $\mathbb{S}(G) = v \in V(G) : d(v) < d_0$.

After this definitions and before entering in technical details, we give an informal summary of the proof. The main idea is to look at the snapshot $G_{\tau 2}$ of $G(\sigma)$. We remind that this is the first graph G_k in the graph process $G(\sigma)$ that has $\delta(G) \geq 2$. We prove that $G_{\tau 2}$ is *whp* hamiltonian. Taking any subset of vertices U in $G_{\tau 2}$, since $\delta(G) \geq 2$, it is reasonable to expect that for every neighborhood of U, its size is at least $2|U|$. So we can infer that $G_{\tau 2}$ is a $(k, 2) - expander$ for $k \sim n$, but this does not gives hamiltonicity yet. An important observation is that since we are proving a hitting time result, we can not add more edges to $G_{\tau 2}$. We must see that the hamilton cycle appears at the same time that the minimum degree becomes greater or equal than 2.

We claim that the snapshot $G_{\tau 2}$ indeed contains a subgraph Γ_0 which is also a $(k, 2) - expander$ but is much more sparse than $G_{\tau 2}$, as it only have a small proportion of its edges. Then, we add boosters of Γ_0 until we finally reach hamiltonicity. The key point is that all graphs that appear in this process of adding boosters are indeed subgraphs of $G_{\tau 2}$. Let us formalize this argumentation, starting with the following lemma:

Lemma 6.7. *[Kri16]* Let $G(\sigma)$ be a random graph process on n vertices. Then, if $G = G_{\tau 2}$, *whp* G has the following properties:

1. $\Delta(G) \leq 10 \log n$ and $\delta(G) \geq 2$.
2. $|\mathbb{S}(G)| \leq n^{0.5}$.
3. G does not contain a path of length at most 4 such that both of its endpoints lie in $\mathbb{S}(G)$.
4. Every vertex subset $U \in [n]$ of size $|U| \leq \frac{n}{\sqrt{\log n}}$ spans at most $|U|(\log n)^{\frac{3}{4}}$.
5. For every pair of disjoint vertex subset U, W of size $|U| \leq \frac{n}{\sqrt{\log n}}$, $|W| \leq |U|(\log n)^{\frac{1}{4}}$, the number of edges of G crossing between U and W is at most $\frac{d_0|U|}{2}$.
6. For every pair of disjoint vertex subsets U, W of size $|U| = |W| = [\frac{n}{\sqrt{\log n}}]$, G has at least $0.5n$ edges between U and W.

Proof. The proofs for every item can be found in [Kri16]. We do not show them here since they consist in standard counts of edges, using the properties of G and $\mathbb{S}(G)$, so in fact they are most mechanical, manipulating binomial coefficients without any specially novel idea. □

Let us take a look at what means every property, informally speaking. The first property gives lower and upper bounds for the degree of every vertex. Notice that $\delta(G) \geq 2$ commes inmediately since $G = G_{\tau 2}$. The second one confirms that $\mathbb{S}(G)$ is in fact a small subset. This is so intuitive because $\mathbb{S}(G)$ is built as the subset of vertices which have much lower degree than the expected degree for a vertex of G. Property three says that vertices in $\mathbb{S}(G)$ are not close *whp*. The last three properties can be summarized as follows: small subsets of vertices spans a small number of edges (P4), small subsets of vertices have not too much edges joining them (P5), and subsets of vertices which are big enough have a big number of edges joining them (P6).

From another point of view, the properties given by this lemma ensure that $G_{\tau 2}$ is a very good expander, but our objective is to look for a subgraph of $G_{\tau 2}$ much more sparse, but good enough as expander. This subgraph Γ_0 is generated as follows. Assume a graph $G = (V, E)$ accomplish the properties of lemma 6.7. We construct Γ_0 choosing edges for every vertex in the graph, but in a different way depending on whether $v \in \mathbb{S}(G)$ or not. Informally speaking, we keep all edges touching $\mathbb{S}(G)$ and we add some random edges. On the one hand, for every $v \in V - \mathbb{S}(G)$, choose a set $E(v)$ of d_0 edges of G incident to v uniformly at random. On the other hand, for every $v \in \mathbb{S}(G)$ we define $E(v)$ as the set of all edges of G touching v. Finally, we define Γ_0 to be the spanning subgraph of G, whose edge set is

$$E(\Gamma_0) = \bigcup_{v \in V} E(v).$$

Let us state a brief property about Γ_0 that we use in a forthcoming proof. As we did with the properties of G, we do not prove it since it is a standard manipulation of binomial coefficients when counting edges.

Lemma 6.8. *[Kri16]* For every pair of disjoint sets U, W of size $|U| = |W| = [\frac{n}{(\log n)^{1/2}}]$, Γ_0 has at least one edge between U and W.

This property is so useful when proving this next lemma. Its purpose is to verify that, once Γ_0 is created, it is indeed a good expander. Notice that is the reason why we Γ_0 is created, so that if it is not a good expander the reasoning of the proof completely fails. The lemma is the following:

Lemma 6.9. *[Kri16]* The subgraph Γ_0 is *whp* (over the choices of $E(v)$) a $(k, 2) - expander$ with at most $d_0 n$ edges, where $k = \frac{n}{4}$.

Proof. By definition of Γ_0, every vertex has degree at most d_0. In fact, has degree exactly equal to d_0 if is not a vertex of $\mathbb{S}(G)$ and can have lower degree if it belongs to $\mathbb{S}(G)$. Since Γ_0 has n vertices, its number of edges must be smaller or equal than $d_0 n$.
Let us now prove that every Γ_0, with $\delta(\Gamma_0) \geq 2$, and satisfying properties of lemmas 6.7 and 6.8, is a $(\frac{n}{4}, 2) - expander$ for every graph G. Remember that it means that, by definition of expander, for every subset S of at most $\frac{n}{4}$ vertices, then $|N_G(S)| \geq 2|S|$.
Let $S \subset [n]$ be a subset of vertices with size $|S| \leq \frac{n}{4}$. Denote $S_1 = S \cap \mathbb{S}(G)$, $S_2 = S - \mathbb{S}(G)$. First, let us suppose that $|S_2| \leq \frac{n}{(\log n)^{1/2}}$. Because of property 3 of lemma 6.7, all vertices from $\mathbb{S}(G)$ are at distance more than 4 from each other, so in particular they are not neighbors. Using also that $\delta(\Gamma_0) \geq 2$, then $N_{\Gamma_0}(S_2)$ (remember that this is the size of the neighborhood of S_1 in $|\Gamma_0|$) is at least $2|S_1|$. By definition, all vertices of S_2 have degree at least d_0 in Γ_0. Using property 4 of 6.7, the set S_2 spans at most $|S_2|(\log n)^{3/4}$ edges in G, and thus in Γ_0. It follows that the number of edges between S_2 and $V - S_2$ is greater or equal than $d_0|S_2| - 2e_{\Gamma_0}(S_2) > \frac{d_0|S_2|}{2}$. Hence, because of property 5 of lemma 6.7, $|N_{\Gamma_0}(S_2)| \geq |S_2|(\log n)^{1/4}$. Finally, again by property 3 of lemma 6.7, the set $S_1 \cup N_{\Gamma_0}(S_1)$ contains only one vertex from $u \cup |N_{\Gamma_0}(u)|$ for every $u \in S_2$. Hence, $|(S_1 \cup N_{\Gamma_0}(S_1)) \cap (S_2 \cup N_{\Gamma_0}(S_2))| \leq |S_2|$.
Then, joining all the results we have got until now, we obtain

$$|N_{\Gamma_0}(S)| = |N_{\Gamma_0}(S_2) - S_1| + |N_{\Gamma_0}(S_1) - (S_2 \cup N_{\Gamma_0}(S_2))|$$

$$\geq |S_2|((\log n)^{1/4} - 1) + 2|S_1| - |S_2| \geq 2(|S_1| + |S_2|) = 2|S|.$$

Let us now prove that in the complementary case, in which $\frac{n}{(\log n)^{1/2}} \leq |S_2| \leq \frac{n}{4}$, the condition for Γ_0 to be an expander is accomplished by S. By lemma 6.8, such S_2 misses at most $\frac{n}{(\log n)}$ vertices in its neighborhood in Γ_0. Furthermore, by property 2 of lemma 6.7, $|S_1| \leq |\mathbb{S}(G)| \leq n^{0.5}$. Altogether,

$$|N_{\Gamma_0}(S)| \geq n - \frac{n}{(\log n)^{1/2}} - |S_2| - |\mathbb{S}(G)| \geq \frac{n}{2} \geq 2|S|,$$

as we desired to prove. □

Now, let us remark that every $\Gamma = (\frac{n}{4}, 2) - expander$ with n vertices is necessarily connected. Let us argue by contradiction to easily see this fact. If Γ is not connected, then consider its connected component C, of size $|C| \leq \frac{n}{2}$, and take U an arbitrary subset of C of size $|U| = min[\frac{n}{4}], |C|$. Then, by definition of expander, $|N_\Gamma U| \geq 2|U|$, and since $2|U| > |C - U|$, this neighborhood falls entirely within C, and this is clearly a contradiction.

Now, let us look at a lemma that allows to add boosters to the subgraph Γ_0, which is a expander but not necessarily Hamiltonian, until we reach Hamiltonicity. This booster must come from within the already existing edges of the random graph, in order to maintain the hitting time. Hence, the following lemma states that *whp* a random graph $G(n, M)$ has a booster with respect to any sparse expander it contains.

Lemma 6.10. *[Kri16]* Let $G(\sigma) = G_0, \dots, G_N$ be a random graph process on n vertices. Denote $G = G_{\tau 2}$. Assume the constant δ_0 is small enough. Then, for every $(\frac{n}{4}, 2) - expander$ $\Gamma \subset G$ with $V(\Gamma) = V(G)$, and $|E(\Gamma)| \leq d_0 n + n$,$\Gamma$ is Hamiltonian, or G contains at least one booster with respect to Γ.

Proof. We claim (we do not prove this result) that every connected $(k, 2) - expander$ Γ is Hamiltonian or has at least $\frac{k^2}{2}$ boosters. So, reasoning by contradiction, let us suppose that a random graph G violates this lemma. It means that G contains a $(\frac{n}{4}, 2) - expander$ which none edge between the $\frac{n^2}{32}$ (or even more) boosters relative to Γ. We know that *whp* $m_1 \leq m \leq m_2$. So let us count the number of edges of Γ. We sum over all the possible values of m in the given interval $[m_1, m_2]$ and we add $o(1)$ regarding at the fact that m can go out of this interval with low probability. Hence, we sum over all possible values i of $|E(\Gamma)|$, and we estimate the number of $(\frac{n}{4}, 2) - expanders$ with i edges in K_n, using $\binom{N}{i}$ as upper bound. Then, we require the edges of Γ to be present in G, but at least $\frac{n^2}{32}$ boosters must not be considered. Finally, we bound the ratio of binomial coefficients using mathematical identities. Thus, we obtain:

$$\sum_{m=m_1}^{m_2} \binom{N}{m}^{-1} \sum_{i \leq d_0 n + n} \binom{N}{i} \binom{N - i - \frac{n^2}{32}}{m - i} + o(1) \leq \sum_{m=m_1}^{m_2} \binom{N}{m}^{-1} \sum_{i \leq d_0 n + n} \binom{N}{i} e^{-\frac{m}{17}} \left(\frac{m}{N}\right)^i + o(1)$$

Now, using that δ_0 is small enough, we can estimate the i-th summand in the previous expression as:

$$\binom{N}{i} e^{\frac{m}{17}} \frac{m^i}{N^i} \leq \left(\frac{eNm}{iN}\right)^i e^{-\frac{m}{17}} \leq \left(\frac{em}{i}\right)^i e^{-\frac{m}{17}} \leq \left(\frac{em}{d_0 n + n}\right)^{d_0 n + n} e^{-\frac{m}{17}} = o(n^{-3})$$

The result follows by summing over all $i \leq d_o n + n$ and over all $m \in [m1, m2]$, since we obtain that all the sum is $o(1)$. It means that *whp* such G which contains a $(\frac{n}{4}, 2) - expander$ with no boosters does not exist, and it concludes the proof. □

Let us now connect all this results to complete the proof of theorem 6.2, this is that *whp* at the moment when the minimum degree of the random graph process becomes 2 it is already

Hamiltonian. By lemma 6.9, *whp* $G_{\tau 2}$ contains an $(\frac{n}{4}, 2)-$*expander* Γ_0 with at most $d_0 n$ edges. We take this Γ_0 and we add boosters until we reach Hamiltonicity. The key fact in this process is this because of lemma 6.10, we can always iterate this process inside $G_{\tau 2}$. So the process of edge addition from Γ_0 always finishes with a subgraph of $G_{\tau 2}$ which is hamiltonian, there is no any point in this process at which we get blocked (i.e. we can not add any booster and our graph is not yet hamiltonian). This fact concludes the proof, since it is sufficient to claim that

$$\tau_2(G(\sigma)) = \tau_H(G(\sigma)).$$

To finish this section, let us talk about a result of hamiltonicity in positional games given by Krivelevich [Kri11] in which expanders are the main tool in the proof. Positional games are a huge field in combinatorics theory, and inside it we focus on *Maker-Breaker games*. This is a triple (H, a, b) where $H = (V, E)$ is an hypergraph whose vertex set is called *board* V and whose edge set E is called *winning sets*. a and b are parameters which play the following role in the game. At each round, the first player, called *Maker*, claims for a free elements in the set V. Then, the second player, called *Breaker*, answers claiming for b elements. We assume that all vertices are free at the beginning, and *Breaker* starts claiming. The game ends when all vertices have been claimed by one of two players. *Maker* wins if and only if he has occupied at least one winning set of E. If not, *Breaker* wins. The main challenge with this kind of problem is if it is possible to know who wins the game respect the parameters a and b.

We focus on the *Hamiltonian game*. In this case, we consider the complete graph K_n, and *Maker*'s goal is to construct a hamiltonian cycle. By construction of the problem, it is necessary to take b larger than a for *Breaker* to win, this is called the bias. The most studied case is taking $a = 1$ and a larger b, which we denote $(1 : b)$. Many results have been achieved for some values of b, in particular the following conjecture was set: The critical bias b for the Hamiltonicity game on K_n is asymptotically equal to $\frac{n}{\log n}$. This conjecture is proved in the mentioned paper [Kri11], thanks to the use of expanders. Indeed, the result obtained is the following:

Theorem 6.11. *Maker* has a strategy to win the $(1 : b)$ Hamiltonicity game played on the edge set of the complete graph K_n on n vertices in at most $14n$ moves, for every for all large enough n, and for every

$$b \leq \left(1 - \frac{30}{\log n^{\frac{1}{4}}}\right) \frac{n}{\log n}$$

This is the final result in this section, but there are also many applications of expanders, either in positional games or not, which are very interesting open areas of research.

7. Conclusions and future work

Random graphs theory is a so powerful tool that allows to achieve strong results in many different fields. However, its main concepts are not so difficult to explain and understand, even to people who has not entered in detail in probability or graph theory. The main objectives in this work were to clearly explain all concepts and to relate it with further research and actual applications in many actual problems. In order to enforce this ideas, let us outline the main ideas and the future work that can be followed from each section.

Our first approach to random graph theory was the first appearance of the concept. It is a fantastic demonstration of how mathematical research works, since random graphs were created as a way to fight completely different problems (combinatorial problems) that could not be solved with classic tools and arguments. Thanks to this beginning, in this first section of the work we present some mathematical concepts that are interesting enough to be the main topic of another work such as Ramsey theory and the probabilistic method.

Then, the following step was to define the random graph model itself. In this work we only worked with the two classic, more common and simpler models, that is $G(n,p)$ and $G(n,M)$. There are more random graph models, increasing its complexity in order to model several phenomena, that can be analogously studied. After introducing $G(n,p)$ and $G(n,M)$, we define graph properties and we set the asymptotic equivalence between both under certain conditions. This is one of the key results in this work, since it allows to work with the model we consider more appropriate in every case, opening a lot of options to face any proof.

Studying graph properties is the main goal of this theory, and that is why threshold functions are essential. They provide conditions on the random graph model to know if this property is accomplished or not. By this way, we can completely characterize a property, meaning that looking at the structure of the random graph it is possible to ensure if the property is fulfilled. Indeed, some of most important properties (and very difficult to study for any general graph) are characterized in this work thanks to random graph and threshold functions.

This characterization is possible thanks to the second key result in this work (theorem 4.2, which states that any monotone property has a threshold). It motivates our study on properties like having subgraphs, since it is a monotone property and can be completely well-known after setting its threshold function. In particular, the study of subgraphs is very important in graph theory, since the presence of some subgraphs provide special structures on the whole graph. The study of the presence of subgraphs can be continued paying more attention to the phase transition, as we did with the appearance of cycles. Moreover, the study of thresholds for some specific not balanced subgraphs is also an interesting field to work out.

The theorem of the birth of the giant component is a classic result that we face using two pretty recent and interesting approaches. They are completely different since the first one is based on a graph search algorithm (*DFS*) and the second one is based on a stochastic process (*Galton-Watson*). What's more, they provide complementary results that are different for each case, which increases the richness of the work. Each method has many different applications, and it can be also interesting see how they are adapted to another kind of problems. Moreover, the result itself is also adapted to other fields (for example, it appears in biological state changes), and the similarities and differences with the result for random graphs are also an exciting topic.

Finally, the last part of the work is about the property of being hamiltonian. This last part is the most recent, since it is a result from 2016. It uses many powerful techniques like hitting times, expanders and boosters. In particular, expanders are a so powerful tool, as we wanted to emphasize with the Hamiltonian game theorem. Applications of expanders, and also positional games, are two interesting topics to do further research. We also introduced the concepts of hypergraphs, which is actually an open field of research with key applications in new technologies like communication networks. Hypergraphs are definitely a perfect topic to continue this work, because of its similarity with random graphs and the wide range of achievements that can be reached by using them.

To conclude this work, I would like to give my personal opinion about it. The bachelor degree thesis was for me a big challenge, and I wanted to create a complete work. With complete, I mean a work that combines classic and recent results, theoretical theorems with practical applications, trying to include as many interesting concepts as possible, but keeping the focus on the main objective. I think that random graphs accomplishes all this conditions, and it is a really interesting topic that merges perfectly all I wanted to do. Of course, the main implicit objectives are to learn as many as possible and understand really well the concepts. On this way, I am very satisfied since I consider I got both objetives.

8. Acknowledgements

I would like to thank my parents, Eva and Miguel Angel, and my brother Pau, for their daily support. I also would like to thank everybody that have attended to me when I was talking about the project. This acknowledgement includes for sure a really big number of friends, classmates and family, since I loved to explain anything to everybody that was, at least, a little bit interested. Last but not least, I would like to thank my advisor, Juan José Rué, for his attitude during the project. His availability to answer any doubt, his ambition to get a rigorous work and his clear and kind way to transmit knowledge have been fundamental for me in order to successfully work this topic.

References

[AKS81] Miklós Ajtai, János Komlós, and Endre Szemerédi. The longest path in a random graph. *Combinatorica*, 1(1):1–12, 1981.

[AS04] Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2004.

[Bol01] Béla Bollobás. *Random graphs*. Number 73. Cambridge university press, 2001.

[CG83] Fan Chung and Charles M Grinstead. A survey of bounds for classical ramsey numbers. *Journal of Graph Theory*, 7(1):25–37, 1983.

[CKK18] Oliver Cooley, Mihyun Kang, and Christoph Koch. The size of the giant high-order component in random hypergraphs. *Random structures & algorithms*, 53(2):238–288, 2018.

[Con09] David Conlon. A new upper bound for diagonal ramsey numbers. *Annals of Mathematics*, pages 941–960, 2009.

[Die05] Reinhard Diestel. Graph theory, vol. 173 of. *Graduate Texts in Mathematics*, page 47, 2005.

[Dir52] Gabriel Andrew Dirac. Some theorems on abstract graphs. *Proceedings of the London Mathematical Society*, 3(1):69–81, 1952.

[ER59] Paul Erdős and Alfréd Rényi. On random graphs. *Publ. Math. Debrecen*, 6:290–297, 1959.

[Erd47] Paul Erdős. Some remarks on the theory of graphs. *Bulletin of the American Mathematical Society*, 53(4):292–294, 1947.

[Erd59] Paul Erdős. Graph theory and probability. *Canadian Journal of Mathematics*, 11:34–38, 1959.

[ES35] Paul Erdős and George Szekeres. A combinatorial problem in geometry. *Compositio mathematica*, 2:463–470, 1935.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[JLR11] Svante Janson, Tomasz Luczak, and Andrzej Rucinski. *Random graphs*, volume 45. John Wiley & Sons, 2011.

[Kri11] Michael Krivelevich. The critical bias for the hamiltonicity game is $(1 + o(1))n/\ln n$. *Journal of the American Mathematical Society*, 24(1):125–131, 2011.

[Kri16] Michael Krivelevich. Long paths and hamiltonicity in random graphs. *Random Graphs, Geometry and Asymptotic Structure*, 84:1, 2016.

[KS13] Michael Krivelevich and Benny Sudakov. The phase transition in random graphs: A simple proof. *Random Structures & Algorithms*, 43(2):131–138, 2013.

[Ore60] Øystein Ore. A note on hamiltonian circuits. *American Mathematical Monthly*, 67:55, 1960.