On random graphs

Leo Davy

Contents

Chapter 1

An introduction to random graphs

1.1 Graph theory

[1] Formally a *graph* G is defined as $G = (V(G), E(G))$, with $V(G)$ designing the vertex set (the points or nodes) of G and $E(G)$ as $E(G) \subseteq \{\{x, y\}, x, y, \in V(G), x \neq y\}$, the set of edges (the lines) of G.

This is the simplest way to define a graph, hence this kind of graph is usually called a *simple graph*. More general graphs can be defined, for instance the loopy graphs, multigraphs, directed graphs or hypergraphs. They are not of major importance in this report so they won't be detailed here, their formal definition can be found in almost any book on graph theory (see for instance [BM08]).

We will call two vertices $u, v \in V(G)$ as *adjacent* if $\{u, v\} \in E(G)$ and we will write it as $u \leftrightarrow v$. We may also refer to edges being adjacent if they share a vertex.

As an example of a graph we can consider the following graph (TODO: ADD GRAPHICAL REPRESENTATION) with $V(G) = \{a, b, c, d, e\}, E(G) = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$ with :

$$e_1 = \{a, b\} \quad e_2 = \{a, c\} \quad e_3 = \{b, c\} \quad e_4 = \{a, d\}$$
$$e_5 = \{c, d\} \quad e_6 = \{a, e\} \quad e_7 = \{c, e\}$$

---

[1] For a very simple introduction to graph theory, see [Tru93], for an advanced review of graph theory, see [BM08]

The *adjacency matrix* of a graph G is the $n \times n$ matrix defined as $A_G = (a_{u,v})$ with $a_{u,v} = \mathbb{1}_{E(G)}(\{u,v\})$

$$
A_G = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array}
\begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\
\left[\begin{array}{ccccc}
0 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0
\end{array} \right] \end{array}
\tag{1.1}
$$

It is interesting to note that an adjacency matrix is real and hermitian, thus all of its eigenvalues are real and the study of their distribution is a classical topic in graph theory.

An important property of graphs is the *degree* of the vertices, so we will denote by $d_G(v)$ the number of edges incident with $v \in V(G)$. Observing that each edge has two ends and that the degree of a vertex is the number of edges having this vertex as an end. We obtain :

$$
\sum_{v \in V(G)} d_G(v) = 2|E(G)|
\tag{1.2}
$$

We also call the degree sequence the non-increasing sequence of its vertex degrees. And we can also define the two following notations that will prove useful in the following of this report, $\delta(G)$ as the minimal degree of G and $\Delta(G)$ as the maximal degree of G.

We define a *path* on a graph as a sequence of edges being two by two adjacent. One of the most fundamental properties of graphs is also the connectivity. We will say that a graph is *connected*, if there is a path connecting any two edges. In fact we will consider simple[2] paths because we are studying simple graphs. It is clear that if a path exists it is always possible to extract a simple path from it. There are two famous kinds of path, the *hamiltonian path*, it is a path that includes all vertices of G. Analogously, an *Eulerian path* is a path in which all edges are used exactly once.

If v is a vertex, we will write $N(v)$ the set of vertices adjacent to v which are called the *neighbours* of v. From this definition we may observe that $d_G(v) = |N(v)|$ if G is a graph without loops. And we will call a (connected) *component* of a vertex the set of vertices that can be reached from this vertex. Then a connected graph is a graph with only one component.

Some interesting graphs to which we will often refer are the complete graph on n vertices, denoted by K_n, and the complete bipartite graph $K_{n,m}$ (TODO : ADD GRAPHICAL REPRESENTATION). A *complete graph* is a graph in which for any vertex, the set of neighbours is the rest of the graph. A graph is *bipartite* if it's set of vertices can be partitioned in two subsets X and Y such that every edge has one end in X and one in Y. The *complete bipartite graph*

---

[2] More generally, the use of the adjective *simple* denotes that we study something without loop or multi-edge

is a bipartite graph such that for all $x \in X$ we have $N(x) = Y$. This implies the same condition on the vertices in Y. We will also call a *cycle*, of size $n \geq 3$, denoted C_n a graph whose vertices can be arranged in a cyclic sequence in such a way that two vertices are adjacent if they are consecutive in the sequence.

As there is usually no confusion possible we will denote $V = V(G), E = E(G), \psi = \psi_G, \dots$

1.2 Random graphs

The study of random graphs is a flourishing area of mathematics since it's founding papers have been published by Erdős and Renyi between 1959 and 1963 [ER59] [ER60] [ER61b] [ER61a] [ER63]). (TODO : REVISE THIS WHOLE SECTION) Since then a lot of work has been done on random graphs, most of the questions on the Erdős-Renyi model have found satisfying answers, and the model being simplistic, many new models have been developed. So we will use the very vague definition by Janson [Sva14].

Definition 1.2.1. A *random graph* is a graph where nodes, or edges, or both are selected by a random procedure.

More formally, a random graph is a probability distribution on a space of graphs. This means that when studying random graphs, we can not only change the distribution but also the space in which we sample.

Usually we make the choice of sampling on a space of graphs containing any graph with a specified number of vertices. [3] Some restrictions on this space might be done, for instance by selecting graphs with each vertex of the same degree (r-regular random graphs), or with the degree sequence following a specific distribution (like the Newman-Strogatz-Watts model [NSW01], [NWS02]). Sometimes random graphs are developped in order to produce a sampling space matching some particular phenomenom observed in real-world networks.

For instance in social networks, two persons having a friend in common are likely to be friends which in the vocabulary of graph theory indicates the presence of a triangle. For this purpose was developed the Watts-Strogatz small world model [WS98] which has a large enough number of triangle. [4] It also features a "small world" characteristic meaning that for most of the vertices there exists a path linking them which is small enough.

1.3 Cayley's formula

In this section we will prove an important result that will be used several times in crucial demonstrations in this report. Indeed it will give an exact result on the

---

[3]This choices differs from what usually appear in the real-world where networks are grown through time, see [Cal+01] for a discussion on fundamental differences between grown random graphs and the static random graphs.

[4]More formally, a positive density of triangles

number of spanning trees, which are the building blocks of connected graphs. A *tree* is a special case of graph structure that can be defined in several equivalent ways. For instance, a tree is a connected graph such that upon removal of any of its edges, it becomes disconnected, equivalently it is a connected and acyclic graph [5]. The adjective *spanning* (on the vertices) meaning that those trees intersect the whole vertex set.

Theorem 1.3.1. Cayley's formula

$$t_n = n^{n-2} \tag{1.3}$$

with t_n the number of spanning trees on n vertices.

The proof will need to make use of directed trees. More generally, we define *directed graphs* as graphs [6] in which the adjacency relation is not symmetric [7]. We also define doubly rooted trees as trees with two special labels "Start" and "End" that can be attached to any vertices. In such a tree, for any vertex, there exists a path to the vertex labelled "End". We will call "SEL" the vertices that are in the path from "Start" to "End". We also denote by DRT_n the set of doubly rooted trees on n vertices.
As a consequence of this definition we have $|DRT_n| = n^2 t_n$. With $|\ |$ denoting the cardinal. To prove the theorem it is then be sufficient to prove that the number of elements in DRT_n is equal to n^n. We will base our approach on Joyal's proof [Joy81] and show a bijection between the set of doubly rooted trees on n vertices and the the set of functions on n elements. The presentation of the proof we use comes from [LZ12].
(TODO : ADD GRAPHICAL REPRESENTATION OF THE PROOF ON SOME EXAMPLE)

Proof. We will use the notation $[n] = \{1, 2, ..., n\}$ and $V = [n]$. Let's take $f : V \longrightarrow V$, and let's consider the graph of f. That is, $\forall v_1, v_2 \in V$ we have $v_1 \to v_2$ if and only if $f(v_1) = v_2$. Drawing such a graph for any function, and will appear two different kind of structures, first there will have directed line leading to cycles, and then cycles. And the whole graph will be a disjoint union of such components. It can be interesting to observe the case in which f is a permutation and then observe that the graph of f is a union of disjoint cycles as expected from the common group theory result.
We now take $C \subseteq V$ the set of vertices that are part of a cycle under the action of f. Equivalently,

$$C = \{x : \exists i \geq 1 s.t. f^i(x) = x\}$$

Let $k = |C|$ and write $C_<$ as $C_< = \{c_1 < c_2 < ... < c_k\}$ the ordered set [8] and now we will construct a graph with the vertex set $D = f(C)$, and the edge set

---

[5]An acyclic graph is a graph that doesn't contain any cycle
[6]Often simply called digraphs (or ditrees)
[7]In a directed graph an edge can only be followed in one direction, we have if T is a ditree, $\forall x, y \in V(T), x \neq y, x \to y \implies y \not\to x$
[8]The c_i being integers we simply order them in increasing order.

$E = \bigcup_{i=1}^{k-1} f(c_i)f(c_{i+1})$. We now have $G = (D, E)$ as a line of k vertices, and we will call $f(c_1)$ the "Start" and $f(c_k)$ the "End" which is an oriented graph. Now we will just append to this line the set of vertices that are not in G. So we construct $\tilde{E} = \bigcup_{x \in VC} xf(x)$ and $\tilde{G} = (V, E \cup \tilde{E})$ is a (directed doubly rooted) tree as it doesn't contain any cycle by construction and is clearly connected. It's obviously directed and doubly rooted. We have now done the biggest part of the proof, that is, going from a function to a doubly rooted tree.

We will now take a doubly rooted tree and transform it in a function. From the definition of trees there is a unique "Start" to "End" (SEL) path.

For vertices on not on the SEL, for instance some vertex j, we define $f(j)$ as the first neighbour on the j to end line.

For vertices on the SEL,

$$SEL = \{x_1, x_2, ..., x_k\}, \text{ and } SEL_< = \{x_{\sigma_1}, x_{\sigma_2}, ..., x_{\sigma_k}\} \tag{1.4}$$

we define $f(x_{\sigma_i}) = x_i, \forall i \in [k]$.

Thus, we have two injective constructions, if composed give the identity, hence we have a bijection between the set of endomorphism of $[n]$ and the space of doubly rooted trees on n vertices. So the proof is complete. $\square$

Chapter 2

The Erdos-Renyi Model

2.1 Different approaches of the same space

As said in the title of the section there are different ways to approach the Erdős-Renyi model that we may call paradigms as they will give us the same kind of results but depending on the context, one might be much more convenient to use than the others.

The first paper published on random graphs that gained notoriety was from Erdős and Renyi in 1959 [ER59], in which they give the following construction :

Definition 2.1.1. We call a random graph $\mathcal{G}_{n,M}$ having n labelled vertices and M edges. That is we choose at random (with equal probability) one of the $\binom{\binom{n}{2}}{M}$ possible graphs.

One may observe that some changes in notations are made between this paraphrasing of the article of Erdős and Renyi, they are made in order to be more adapted with the modern study of random graphs. We will also adopt for the following $N = \binom{n}{2}$ to denote the total number of edges possible on n labelled vertices.

We then arrive to our main model that has been the most extensively studied in the literature of random graphs, that is $\mathcal{G}_{n,p}$ on which the coin tosses are no longer necessarily fair, but the probability of drawing an edge is now p. This model was first introduced by Gilbert in 1959, see [Gil59][1]. And the coin tosses are still independent. Now if we denote by e_G the number of edges of a graph G on the vertex set $[n]$. We have :

$$\mathbb{P}(G) = p^{e_G}(1-p)^{N-e_G} \tag{2.1}$$

This model is called the *binomial model*. It is easily seen that this model is asymptotically equivalent to the first one if Np is close to M on several aspects.

---

[1]It is interesting to note that Gilbert's paper who was at the time working at Bell's laboratories formulated the question of connectivity based on real life phone networks.

The third model that we will investigate is on the form of a Markov process, see in Annex for a discussion on properties used here from Markov chains. At time 0 there is no edge and an edge is selected at random among all of the possible edges. At time t, the edge is chosen among all the edges not already present in the graph. We denote this process by $\{\mathcal{G}_{n,t}\}_t$, with t the number of edges added. It is clear that this model is perfectly equivalent to the first model presented in the case $t = M$. This model was also introduced in 1959 by Erdős and Renyi and is usually referred to as the *random graph process*. The advantage of this model is that it allows one to study properties on the verge of their realisations. For instance, using this model Bollobás and Thomason [BT85] proved that a graph is fully connected, when the last connection made is between an isolated vertices and the giant component. Equivalently, when $\delta(G) > 1$. But we will study this in the following.

The three models introduced above ($\mathcal{G}_{n,M}$, $\mathcal{G}_{n,p}$ and $\{\mathcal{G}_{n,t}\}_t$) are usually designated in the literature as Erdős-Rényi model. However, even if the first two are quite similar (static random graphs) the last one is of a different nature as it is dynamic.

In this report we will mainly focus on $\mathcal{G}_{n,p}$ which is in fact easier to study than $\mathcal{G}_{n,M}$. [2] We will see in (2.3.2) that there is a strong relation between those two models.

2.2   Connectivity

One of the most fundamental structure of a graph will be its number of components. Hence, the first question we will try to ask is with which probability a random graph in the Erdős-Renyi model is connected. It is essential to answer this question as many other questions might not make sense or have an obvious answer on a graph that is not connected (for instance the diameter, the existence of Hamiltonian paths or the stability number of the graph).

It is also an interesting first topic to have an insight of the kind of elegant results that arise from the study of random graphs. The main aim of this section will be to prove the following theorem in a detailed way as it is the first random graphs proof that we will study.

Theorem 2.2.1. Let $p = p(n) = \frac{\log(n)+c}{n}$, $c \in \mathbb{R}$
Then $\lim_{n \to \infty} P(G \in \mathcal{G}_{n,p}$ is connected $) = e^{-e^{-c}}$

The proof of this theorem will be in two parts, first we will show that a graph will be connected if and only if there are no isolated vertices and then we will estimate the distribution that follows the number of isolated vertices. We designate by giant component a component with a size increasing linearly in n.

Theorem 2.2.2. With $p = \frac{\log(n)+c}{n}$ Almost every $G \in \mathcal{G}_{n,p}$ consists of a single giant component and isolated vertices.

---

[2]There is usually no confusion possible on the model being studied.

Proof. This proof can be found in [Spe14] or [Bol01].

During this proof we will consider the random value X_k that counts the number of vertices that are in components of size k. So, let's estimate the probability $P(X_2 > 0) = P(X_2 \geq 1)$. In order to do so we will use the method of first moment.

$$\mathbb{P}(X_2 \geq 1) \leq \mathbb{E}(X_2) = \binom{n}{2}\mathbb{P}(\text{"drawing an isolated edge"}) \tag{2.2}$$

$$= \binom{n}{2}p((1-p)^{n-1})^2 \tag{2.3}$$

$$\leq (\frac{ne}{2})^2 p(e^{-p})^{2(n-2)} \tag{2.4}$$

$$= \mathcal{O}\left(n^2 p \frac{e^2 e^{-2p(n+1)}}{4}\right) \tag{2.5}$$

$$= \mathcal{O}(n^2 p) \tag{2.6}$$

So it's sufficient that $p = o(n^{-2})$ in order to have almost surely no edges in G. This is clearly satisfied by the p we use in the theorem.

However, this is not sufficient to prove that there is no isolated component other than isolated vertices. We will observe that there can't have any component of size larger than $\lceil\frac{n}{2}\rceil$ that is not the largest component in the graph. Hence, we will study the probability that there is any component of intermediary size that is not the greatest component.

$$\mathbb{P}(X_k \geq 1) \leq \mathbb{E}(X_k) \quad , \forall k \geq 3 \tag{2.7}$$

$$\leq \binom{n}{k}k^{k-2}q_k \tag{2.8}$$

$$\leq \binom{n}{k}k^{k-2}p^{k-1}((1-p)^{n-k})^k \tag{2.9}$$

In the above, q_k represents the probability that a spanning tree on k vertices doesn't connect to the greatest connected components, the total number of spanning trees on k-vertices being k^{k-2} from Cayley's formula (1.3.1). A tree on k vertices having $k-1$ edges, this means in terms of probability that it must have $k-1$ "success" and on each of the k vertices $n-k$ failures. Which leads to the following line.

Now we will try to have an upper bound of the RHS such that the sum on k will converges to a $o(n^{-\delta})$ for some $\delta > 0$.

$$\mathbb{P}(X_k \geq 1) \leq k^{-2}p^{-1}(\frac{ne}{k})^k k^k p^k e^{-p\frac{n}{2}(n-k))} \tag{2.10}$$

$$\leq k^{-2}p^{-1}((\frac{ne}{k})kpe^{-p\frac{n}{2}})^k \tag{2.11}$$

$$\leq p^{-1}(nepe^{-p\frac{n}{2}})^k \tag{2.12}$$

If we denote the bracketed term by A, then

$$A = \mathcal{O}(\log(n)n^{-\frac{1}{2}}) \tag{2.13}$$

Hence, A goes to 0 without any condition on k, so we obtain

$$\sum_{k=3}^{\lceil n/2 \rceil} \mathbb{P}(X_k \geq 1) \leq p^{-1} \sum_{k=3}^{\lceil n/2 \rceil} A^k = o(1) \tag{2.14}$$

This gives the fact that whether $\mathcal{G}_{n,p}$ is connected doesn't depend on the existence of isolated connected components of size 2 or more. So it only depends on the existence of isolated vertices. So we have proved (2.2.2). $\qquad\square$

The proof of (2.2.1) can be done using the second moment method or Stein's method, see [JŁR00]. (TODO : CHOOSE WHICH ONE TO PRESENT.).

2.3 Existence of thresholds

[3] One of the most surprising features on random graphs, which seems to have motivated Erdős to publish his results from 1959 (TODO : FIND THE REF THAT SAID THAT...), is the existence of thresholds. Indeed the property of appearance of certain graph properties will be either 0 or 1 for a great range of functions p.

For instance from the previous theorem we observe that $\forall \epsilon > 0$ if $p = (1 + \epsilon)\frac{\log(n)}{n}$ or if $p = \frac{\log(n) + \omega(n)}{n}$ with $\lim_{n\to\infty} \omega(n) = \infty$ then $\lim_{n\to\infty} \mathbb{P}(G \in \mathcal{G}_{n,p}$ is connected$) = 1$. We would say that for such p, $\mathcal{G}_{n,p}$ is almost surely connected. We also observe that if in the previous definitions of p we changed the $+$ signs into $-$ signs, we would find that $\mathcal{G}_{n,p}$ is almost surely disconnected. Hence, the function p from (2.2.1) has some very peculiar behaviour on the property of connectivity. We will call such a function a *threshold* (here, for connectivity).

The aim of this section is to detail formally what we mean by a graph property and a threshold function. We will also show a formal relation between $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,M}$ and a proof of the existence of thresholds on a family of graph properties. It has been shown by Bollobás and Thomason [BT87] that this is in fact not exclusive to random graphs, but true for all monotone properties on random subsets.

Definition 2.3.1. We will call a *graph property* a family of graphs that is closed under isomorphism.

This means that a graph property is independent of the labelling and of the drawing of the graph. We can refine properties in the following definition.

Definition 2.3.2. A property is *monotone* [4] *increasing* (resp. *decreasing*) if it's stable under the the addition (resp. removal) of an edge. A graph property $\mathcal{Q}$ is *convex* if when $A, C \in \mathcal{Q}$ and $A \subseteq B \subseteq C$[5] then $B \in \mathcal{Q}$.

---

[3] The proofs and results from this section are from [JŁR00] and [Bol01]

[4] A property is monotone if it is either increasing or decreasing

[5] $\subseteq$ is the inclusion of the edges on the same set of vertices

For instance, being connected or containing a specific subgraph are monotone increasing properties whereas being planar or containing an isolated vertex are monotone decreasing. As an example of property that is neither monotone increasing or decreasing, we can think of being k-regular for some k (this means that all vertices are of degree k). Having exactly k isolated vertices is an example of a convex not monotone property.

Here is a theorem showing that monotone increasing properties make probability distributions on these properties also monotone increasing.

Theorem 2.3.1. Suppose $\mathcal{Q}$ is a monotone increasing property and $0 \leq M_1 \leq M_2 \leq N$ and $0 \leq p_1 \leq p_2 \leq 1$.
Then
$$\mathbb{P}_{M_1}(\mathcal{Q}) \leq \mathbb{P}_{M_2}(\mathcal{Q}) \text{ and } \mathbb{P}_{p_1}(\mathcal{Q}) \leq \mathbb{P}_{p_2}(\mathcal{Q})$$

Proof. The first inequality is clear, as the only difference between the two spaces on which we evaluate the property $\mathcal{Q}$ is that on the RHS edges have been added, hence, the probability of realising a monotone increasing has been increased.
For the second inequality, let $p = \frac{p_2 - p_1}{1 - p_1}$. Let $G_1 \in \mathcal{G}_{n,p_1}, G_2 \in \mathcal{G}_{n,p}$
So if $G_2 = G_1 \cup G$ it's edges are chosen with probability $p_1 + p - p_1 p = p_2$. So G_1 is in G_2, the property being monotone increasing, we have $\mathbb{P}_{p_1}(\mathcal{Q}) \leq \mathbb{P}_{p_2}(\mathcal{Q})$
$\square$

The following result follows from definition with $\mathcal{Q}$ a monotone increasing property
$$\mathbb{P}(\mathcal{Q}) = \sum_{A \in \mathcal{Q}} p^{|A|}(1-p)^{N-|A|} \tag{2.15}$$

However this result requires to know all of the elements in $\mathcal{Q}$ and as we are often interested with properties for very large n this result won't be magical...
The following theorem shows that if we know quite accurately $\mathbb{P}_M(\mathcal{Q})$ for every M close to pN then we know $\mathbb{P}_p(\mathcal{Q})$ with a comparable accuracy. The converse being clearly false, for instance the property of containing M edges.

Theorem 2.3.2. Suppose $\mathcal{Q}$ is any property and $0 < p = M/N < 1$
Then $\mathbb{P}_M(\mathcal{Q}) \leq 3\sqrt{M}\mathbb{P}_p(\mathcal{Q})$

Proof. Let $\mathcal{Q}$ be any property, then we will write $\mathcal{Q}$ as a partition based on the number of edges in each graph contained in $\mathcal{Q}$.
So we have
$$\mathcal{Q} = \bigsqcup_{m=0}^{N} \mathcal{Q}_m \quad , \text{ with } \forall G \in \mathcal{Q}_m, e(G) = m$$

$\mathbb{P}_m(\mathcal{Q}) = |\mathcal{Q}_m| \binom{N}{M}^{-1}$ From this we can obtain, with $q = 1 - p$

$$\begin{aligned}
\mathbb{P}_p(\mathcal{Q}) &= \sum_{A \in \mathcal{Q}} p^{|A|} q^{N-|A|} = \sum_{m=0}^{N} \sum_{A \in \mathcal{Q}_m} p^{|A|} q^{N-|A|} \\
&= \sum_{m=0}^{N} \sum_{A \in \mathcal{Q}_m} p^m q^{N-m} = \sum_{m=0}^{N} |\mathcal{Q}_m| p^m q^{N-m} \\
&= \sum_{m=0}^{N} p^m q^{N-m} \binom{N}{M} \mathbb{P}_m(\mathcal{Q}) \geq \binom{N}{M} p^M q^{N-M} \mathbb{P}_M(\mathcal{Q}) \\
&\geq \mathbb{P}_M(\mathcal{Q})(e^{\frac{1}{6M}} \sqrt{2\pi pqN})^{-1}
\end{aligned}$$

So we have (A.2.2) (TODO : PROVE LAST INEQ IN ANNEX OR ADD REF)

$$\mathbb{P}_M(\mathcal{Q}) \leq \mathbb{P}_p(\mathcal{Q}) e^{\frac{1}{6M}} \sqrt{2\pi qM} \tag{2.16}$$

Observing that $q \leq 1$ and $\sqrt{2\pi} e^{\frac{1}{6}} \approx 2.961... < 3$ the proof is complete. $\qquad\square$

The previous section was about connectivity in $\mathcal{G}_{n,p}$, in this section we have seen that connectivity can be characterised as a monotone increasing property. Also it was observed that the function p was somehow best possible, by that we mean that modifying it slightly would imply to only have a zero-one law. We call such a function p a threshold (in that case for the connectivity).

More formally, let $\mathcal{Q}$ a monotone increasing property, in $\mathcal{G}_{n,p}$, we call $\hat{p} = \hat{p}(n)$ a threshold if

$$\mathbb{P}(\mathcal{G}_{n,p} \in \mathcal{Q}) \to \begin{cases} 0 & \text{if } p/\hat{p} \to 0, \\ 1 & \text{if } p/\hat{p} \to \infty. \end{cases} \tag{2.17}$$

Analogously, in $\mathcal{G}_{n,M}$, we call $\hat{M} = \hat{M}(n)$ a threshold if

$$\mathbb{P}(\mathcal{G}_{n,M} \in \mathcal{Q}) \to \begin{cases} 0 & \text{if } M/\hat{M} \to 0, \\ 1 & \text{if } M/\hat{M} \to \infty. \end{cases} \tag{2.18}$$

In fact, thresholds are unique with respect to the multiplication by a positive constant. So for the following, we should denote a threshold for a property as *the* threshold.

As said in the introduction of this section, the fact that thresholds could be found for many of the properties that where investigated is one of the main reason behind the study of random graphs. In fact, the following theorem from Bollobás and Thomason (TODO : ADD REF)confirms that we can always expect the existence of a threshold if the property investigated is non trivial.

Theorem 2.3.3. Every non-trivial monotone graph property has a threshold

Proof. We consider without loss of generality that $\mathcal{P}$ is a monotone increasing graph property. Given $0 < \epsilon < 1$ we define $p : [0,1] \to [0,1]$ such that :

$$\mathbb{P}(\mathcal{G}_{n,p} \in \mathcal{P}) = \epsilon \tag{2.19}$$

The existence of p is guaranteed from (2.15) because it is an increasing polynomial in p, from 0 to 1. Indeed, we know that it is increasing from (2.3.1).
We will show that $p^* = p(\frac{1}{2})$ is a threshold for $\mathcal{P}$ through a coupling argument. We take $G_1, G_2, ..., G_k$ independent random variables in $\mathcal{G}_{n,p}$. Then we claim that $G_1 \cup G_2 \cup ... \cup G_k$ is distributed as a random variable in $\mathcal{G}_{n,1-(1-p)^k}$. Which is clear when we observe the following equivalence for $p_1 < p$ and p_2.

$$1 - p = (1 - p_1)(1 - p_2) \iff p = p_1 + p_2 - p_1 p_2 \tag{2.20}$$

The previous equation being satisfied with $p = 1 - (1-p)^k$, $p_1 = 1 - (1-p)^{k-1}$ and $p_2 = p$.
We prove the claim by induction, the initialisation is clearly true, then if we write $G = G_1 \cup ... \cup G_{k-1}$.

$$\mathbb{P}(G \cup G_k = H) = \mathbb{P}(G = H) + \mathbb{P}(G_k = H) - \mathbb{P}(G \cap G_k = H) \tag{2.21}$$
$$= \mathbb{P}(G = H) + \mathbb{P}(G_k = H) - \mathbb{P}(G = H)\mathbb{P}(G_k = H) \tag{2.22}$$

And we can conclude in the same way as in (2.3.1). We may now use Bernoulli's inequality (A.1.4), $1 - (1-p)^k \leq kp$ to obtain

$$\mathbb{G}_{n,1-(1-p)^k)} \subseteq \mathbb{G}_{n,kp} \tag{2.23}$$

and so $\mathbb{G}_{n,kp} \notin \mathcal{P}$ implies $G_1, G_2, ..., G_k \notin \mathcal{P}$. We obtain

$$\mathbb{P}(\mathbb{G}_{n,kp} \notin \mathcal{P}) \leq (\mathbb{P}(\mathbb{G}_{n,p} \notin \mathcal{P}))^k \tag{2.24}$$

Let ω be a function of n growing arbitrarily slowly such that $\lim_{n\to\infty} \omega(n) = \infty$. Suppose also $p^* = p(\frac{1}{2})$ and $k = \omega$, then

$$\mathbb{P}(\mathbb{G}_{n,\omega p^*} \notin \mathcal{P}) \leq 2^{-\omega} = o(1) \tag{2.25}$$

On the other hand,

$$\frac{1}{2} = \mathbb{P}(\mathbb{G}_{n,p^*} \notin \mathcal{P}) \leq (\mathbb{P}(\mathbb{G}_{n,\frac{p^*}{\omega}} \notin \mathcal{P}))^{\omega} \tag{2.26}$$

Finally,

$$\mathbb{P}(\mathbb{G}_{n,\frac{p^*}{\omega}} \notin \mathcal{P}) \geq 2^{-\frac{1}{\omega}} = 1 - o(1) \tag{2.27}$$

This proves that p^* is a threshold for $\mathcal{P}$. $\qquad\qquad\square$

2.4   The stability number

[6] Another property of graphs that one might be interested to study is the stability number. The stability number of a graph is the size of the largest set of vertices we can choose in a graph such that no two vertices are adjacent. One of the reasons that makes this an interesting property to study is that it is linked to one of the most fundamental property of graphs, $\chi(G)$, the *chromatic number*. Indeed, the chromatic number being the smallest number of colours[7] that makes a *proper colouring* of a graph. That means a colouring on which there are no two adjacent vertices of the same colour. We are then led to a similar question than finding the stability number of a graph. If we denote by $\alpha(G)$ the stability number of a graph, we have the simple lower bound [8]

$$\chi(G) \geq \frac{n}{\alpha(G)} \tag{2.28}$$

Before giving a lower bound on the stability number of a graph it might be interesting to denote that the notion of stable set is dual to the notion of clique and is analogous to the notion of perfect matching that concerns the edges. Although the following theorem will give a bound that is quite tight in $\mathcal{G}_{n,p}$ the problem of finding the actual maximum stable set of a graph is a NP-hard problem.

Theorem 2.4.1. The stability number of a graph in $\mathcal{G}_{n,p}$, is at most $\lceil 2p^{-1} \log(n) \rceil$.

Proof. Let $G \in \mathcal{G}_{n,p}$ and $S \subseteq V$ such that V contains $k + 1$ vertices. Then we have

$$\mathcal{P}(\text{"S is a stable set"}) = (1 - p)^{\binom{k+1}{2}} \tag{2.29}$$

as none of the $\binom{k+1}{2}$ possible edges must be selected.
Let's define our random values as follow, $X_S = \mathbb{1}(\text{"} S \text{ is a stable set "})$ and

$$X_{k+1} = \sum_{\substack{S \subseteq V \\ |S|=k+1}} X_S \tag{2.30}$$

the random variable counting the number of stable sets of size $k + 1$, so what we are investigating here is the rank α such that $X_k = 0, \forall k > \alpha$. Such an α

---

[6]This section is from [BM08]

[7]A colouring of a graph is just assigning to each edge a colour. Colours can be thought as "red, green, blue, ..." or as numbers.

[8] We use this link with the chromatic as a simple motivation for studying the stability number. It is a number which is in fact of great importance in graph theory but on matters that are out of the scope of this report. See for instance Chapter 12 in [BM08].

would then be maximal stability number.

$$\mathbb{E}X_{k+1} = \sum_{\substack{S \subseteq V \\ |S|=k+1}} \mathbb{E}X_S = \sum_{\substack{S \subseteq V \\ |S|=k+1}} \mathbb{P}(X_S = 1) \tag{2.31}$$

$$= \sum_{\substack{S \subseteq V \\ |S|=k+1}} (1-p)^{\binom{k+1}{2}} = (1-p)^{\binom{k+1}{2}} \sum_{\substack{S \subseteq V \\ |S|=k+1}} 1 \tag{2.32}$$

$$= \binom{n}{k+1}(1-p)^{\binom{k+1}{2}} \tag{2.33}$$

Now we will use the common inequalities $\binom{n}{k+1} \leq \frac{n^{k+1}}{(k+1)!}$ and $(1-p) \leq e^{-p}$. And we have

$$\mathbb{E}X_{k+1} \leq \frac{n^{k+1}}{(k+1)!} e^{-p\binom{k+1}{2}} = \frac{n^{k+1}}{(k+1)!} e^{-p\frac{k(k+1)}{2}} \tag{2.34}$$

$$\leq \frac{(ne^{-p\frac{k}{2}})^{k+1}}{(k+1)!} \tag{2.35}$$

So, if we consider $k = \lceil 2p^{-1}\log(n) \rceil \leq 2p^{-1}\log(n)$ we have that $ne^{-p\frac{k}{2}} \leq 1$. We finally obtain

$$\mathbb{E}X_{k+1} \longrightarrow_{n\to\infty} 0 \tag{2.36}$$

And then $X_{k+1} = 0$ almost surely which proves the theorem.

$\square$

2.5 The diameter

[9]

Definition 2.5.1. The *diameter* of a graph is the greatest distance between any pair of vertices. We denote it by $\mathrm{diam}(G)$ and say it is equal to ∞ if the graph is not connected.

It is quite easy to understand that the diameter is a value that is a great importance particularly in applied systems. For instance, the small world phenomena is quite notorious (and will be discussed later in this report) but we can also think of optimisation problems in which the fact that two points are far apart might be of great consequences. This section won't be focused on real world applications of the diameter because as we will see we are studying a model which is not realistic for that. Hence we will first discuss some graph theoretic problems and results on the diameter and after giving the main theorem on the diameter we will demonstrate it through several technical lemmas, some of which will be admitted. Finally some corollary will be obtained from the theorem.

---

[9]This section uses [Bol01]

One of the challenging questions in graph theory is estimating the following function

$$n(D, \Delta) = max\{|G|, \operatorname{diam}(G) \leq D, \Delta(G) \leq \Delta\} \tag{2.37}$$

n is the function that for a fixed diameter D and a fixed maximal degree Δ, gives the graph with a maximal number of vertices that verifies both conditions. This is the kind of problem that is part of *extremal graph theory*.
For instance if we take $\Delta = 2$ we obtain easily by construction that a graph that maximises the number of vertices with a diameter D is a $(2D + 1)$-cycle. Hence,

$$n(D, 2) = 2D + 1, \forall D \in \mathbb{N} \tag{2.38}$$

But it is in fact very hard to obtain such a formula for other values of D or Δ. If we take a graph of max degree Δ we observe that there are at most $\Delta(\Delta-1)^{k-1}$ vertices at distance k from a chosen vertex. It is easily to be convinced of this simply by drawing such a graph. From this very simple construction we can obtain the following upper bound

$$n(D, \Delta) \leq 1 + \Delta \sum_{j=1}^{D} (\Delta - 1)^{j-1} = \frac{\Delta(\Delta - 1)^D - 2}{\Delta - 2} = n_0(D, \Delta) \tag{2.39}$$

n_0 is called the Moore bound and a graph for which the Moore bound is best possible is called a Moore graph. The *Petersen's graph* is such a graph with parameter $D = 2$ and $\Delta = 3$. (TODO : ADD A DRAWING OF PETERSEN'S GRAPH).
When $D = 2$ explicit constructions of Moore graphs have been found with $\Delta = 2$ (pentagon) and $\Delta = 3$ (Petersen's graph) as we have seen above. There also exists an explicit construction with $\Delta = 7$, the Hoffman-Singleton graph on 50 vertices. There might exist a graph with $\Delta = 57$ on 3250 vertices but its existence is still an open-question. However, it is proved that no other graph of diameter 2 can be a Moore graph, see Hoffman Singleton (TODO: ADD REF).
In the end of this section we will give a quite good lower bound on n, in order to do so, we will see that we can select a probability function p such that the value of the diameter is highly concentrated.
In the following we will consider that $d = d(n)$ is a natural number (representing the diameter) and c is a positive real number.
Now let's give some functions that will allow us to make a study of the diameter in random graphs. If we choose x a vertex in a graph, then we define $\Gamma_k(x)$ as the set of vertices at distance k from x. And from this we will define $N_k(x)$ as the set of vertices of distance less than or equal to k. Formally, we have

$$\Gamma_k(x) = \{v : d(x, v) = k\} \tag{2.40}$$

$$N_k(x) = \bigcup_{i=1}^{k} \Gamma_i(x) \tag{2.41}$$

And we can link those with the diameter using, $\operatorname{diam}(G) \leq d$ if and only if $N_d(x) = V(G), \forall x$.

Similarly $\operatorname{diam}(G) \geq d$ if and only if $\exists y, N_{d-1}(y) \neq V(G)$.

Now we will define the probability function that we will use in the following of this section as

$$p^d n^{d-1} = \log(\frac{n^2}{c}) \quad , \text{ for some } c > 0 \tag{2.42}$$

In order for $d = d(n)$ to be finite we need [10]

$$\frac{pn}{(\log(n))^3} \longrightarrow \infty \tag{2.43}$$

The aim of this section will be to prove the following theorem.

Theorem 2.5.1. Using all previous definitions on p, d and c. We have,

$$\mathbb{P}(\operatorname{diam}(\mathbb{G}_{n,p}) = d) \longrightarrow e^{-\frac{c}{2}} \tag{2.44}$$

$$\mathbb{P}(\operatorname{diam}(\mathbb{G}_{n,p}) = d + 1) \longrightarrow 1 - e^{-\frac{c}{2}} \tag{2.45}$$

This theorem states that in $\mathcal{G}_{n,p}$ with p defined as in (2.42) the diameter is spread on only two values. As a corollary we clearly have

Corollary 2.5.1.1. Using all previous definitions on p, d and c. We have,

$$\mathbb{P}(\operatorname{diam}(\mathbb{G}_{n,p} \in \{d, d+1\}) \longrightarrow 1 \tag{2.46}$$

In fact the number of values that the diameter can take has been fully resolved by Chung and Lu in [CL01].

Before proving such a theorem we will need some technical lemmas and assumptions. So we will give here some equations for reference later. Also

$$p(\frac{\log(n)}{n})^{-1} \longrightarrow_{n \to \infty} \infty \tag{2.47}$$

that should not be too surprising as the threshold for connectivity is $\frac{\log(n)+c}{n}$ as shown before. We may also consider that $p = o(n^{-\frac{1}{2}+\epsilon}), \forall \epsilon > 0$ as it can be shown that otherwise the diameter of $\mathcal{G}_{n,p}$ is lower or equal than 2. One might be interested in what $p(n, d, c)$ or $d(n, p, c)$ look like so here they are

$$p = n^{\frac{1}{d}-1}(\log(\frac{n^2}{c}))^{\frac{1}{d}} \tag{2.48}$$

$$d = \frac{1}{\log(pn)}(\log(n) + \log\log(n) + log(2) + \mathcal{O}(\frac{1}{\log(n)})) \tag{2.49}$$

$$= \mathcal{O}((1 + o(1))\frac{\log(n)}{\log\log(n)}) \tag{2.50}$$

And finally

$$p(pn)^{d-2} = o(1) \tag{2.51}$$

---

[10] See (2.49)

The first lemma that we present here gives a tail inequality of $\Gamma_k(x)$ conditionally on some space that we will now define. $\Omega_k \subseteq \mathcal{G}_{n,p}$ is a set of graphs with $a = |\Gamma_{k-1}(x)|$ and $b = |N_{k-1}(x)|$ that satisfy

$$\begin{cases} \frac{1}{2}(pn)^{k-1} \le a & \le \frac{1}{2}(pn)^{k-1} \\ b & \le 2(pn)^{k-1} \end{cases} \tag{2.52}$$

Lemma 2.5.2. Let x be a fixed vertex.
$1 \le k = k(n) \le d - 1$
And $K = K(n)$ that satisfy $6 \le K \le \frac{1}{12}\sqrt{pn\frac{1}{\log(n)}}$ We also define

$$\alpha_k = K\sqrt{\frac{\log n}{(pn)^k}}, \quad \beta_k = p(pn)^{k-1}, \quad \gamma_k = 2\frac{(pn)^{k-1}}{n} = \frac{2\beta_k}{pn} \tag{2.53}$$

Then

$$\mathbb{P}(|\gamma_k(x) - apn| \ge (\alpha_k + \beta_k + \gamma_k)apn \quad | \quad \Omega_k) \le n^{-\frac{K^2}{9}} \tag{2.54}$$

Proof. Knowing that we are in the space Ω_k, then the probability that some vertex y is at distance k, that is the probability that y is not in $N_{k-1}(x)$ but is connected to $\Gamma_{k-1}(x)$ is

$$p^a = 1 - (1 - p)^a \tag{2.55}$$

Hence, the random value $|\Gamma_k(x)|$ follows a binomial on $n - b$ elements, with a probability p_a. In the following inequalities we will assume that n is large enough, this will allow us to do certain inequalities that are not really precise but asymptotically satisfying. Also to make it more easy for the reader we consider that all of these inequalities take place in Ω_k without writing it explicitly.

$$|\gamma_k(x) - apn| \ge (\alpha_k + \beta_k + \gamma_k)apn \tag{2.56}$$
$$\ge (\alpha_k + \beta_k)apn + \gamma_k apn \tag{2.57}$$
$$\ge (\alpha_k + \beta_k)apn + ap(n - n_k) \tag{2.58}$$
$$\ge (\alpha_k + \beta_k + 1 - \frac{n_k}{n})apn \tag{2.59}$$
$$\ge (\alpha_k + \beta_k)apn \ge (\alpha_k + \beta_k)apn_k \tag{2.60}$$
$$\tag{2.61}$$

From this first sequence of inequalities we managed to remove γ_k and we used some n_k that is less than n. The point in evaluating these inequalities is that we now have

$$\mathbb{P}(|\gamma_k(x) - apn| \ge (\alpha_k + \beta_k + \gamma_k)apn \quad | \quad \Omega_k) \tag{2.62}$$
$$\le \mathbb{P}(|\gamma_k(x) - apn_k| \ge (\alpha_k + \beta_k)apn_k \quad | \quad \Omega_k) \tag{2.63}$$

Now using $ap - p_a \le \beta_k ap$ and the triangular inequality we have

$$\le \mathbb{P}(|\gamma_k(x) - p_a n_k| \ge \alpha_k apn_k \quad | \quad \Omega_k) \tag{2.64}$$
$$\le \mathbb{P}(|\gamma_k(x) - p_a n_k| \ge \alpha_k p_a n_k \quad | \quad \Omega_k) \tag{2.65}$$
$$\tag{2.66}$$

18

And using Theorem 1.7 (A.2.2) from Bollobás (TODO : ADD IT IN THE AN-NEX AND PROVE IT, JUST A TAIL INEQ ON BINOM)we have

$$\leq \frac{1}{\sqrt{\alpha_k^2 p_a n_k}} \exp(-\frac{1}{3}\alpha_k^2 p_a n_k) \tag{2.67}$$

$$\leq \exp(-\frac{1}{3}\alpha_k^2 p_a n_k) \tag{2.68}$$

And using $p_a \geq pa(1 - \frac{pa}{2})$, $a \geq \frac{1}{2}(pn)^{k-1}$ and using $n_k = n - b$ we obtain $p_a n_k > \frac{(pn)^k}{3}$ that we insert in the previous inequality that will give us the result expected.

$$\leq \exp(-\alpha_k^2 \frac{(pn)^k}{9}) = n^{-\frac{K^2}{9}} \tag{2.69}$$

$\square$

We will now prove another lemma

Lemma 2.5.3. Let $K > 12$ a constant and $\alpha_k, \beta_k, \gamma_k, k = 1, ..., d-1$ as before. Set

$$\delta_k = \exp(2\sum_{l=1}^{k}(\alpha_l + \beta_l + \gamma_l)) - 1 \tag{2.70}$$

Then if n is sufficiently large, with probability at least $1 - n^{-K-2}$ for every vertex x and every natural number $k, 1 \leq k \leq d-1$ we have

$$||\Gamma_k(x)| - (pn)^k| \leq \delta_k(pn)^k \tag{2.71}$$

Proof. As $\delta_{d-1} \longrightarrow_{n\to\infty} 0$ we may assume that $\delta_{d-1} < \frac{1}{4}$ For a fixed vertex x, we denote by Ω_k^* the set of graph for which

$$||\Gamma_l(x) - (pn)^l| \leq \delta_l(pn)^l \quad , 0 \leq l \leq k \tag{2.72}$$

And it is easy to verify that it is decreasing. Anf if one replaces $|\Gamma_l(x)|$ by a it is clear that we have

$$\Omega_k^* \subseteq \Omega_{k-1}^* \subseteq \Omega_k \tag{2.73}$$

Now, simply applying Bayes formula (for the complementary) we have

$$1 - \mathbb{P}(\Omega_k^*) = 1 - \mathbb{P}(\Omega_{k-1}^*) + \mathbb{P}(\Omega_{k-1}^*)\mathbb{P}(||\Gamma_k(x)| - (pn)^k| \leq \delta_k(pn)^k|\Omega_{k-1}^*) \tag{2.74}$$

If $G \in \Omega_{k-1}^*$ and $|a| = |\Gamma_{k-1}(x)|$ and by applying the definition of belonging to Ω_{k-1}^* and multiplying both sides by pn we have

$$|(pn)^k - apn| \leq \delta_{k-1}(pn)^k \tag{2.75}$$

And we obtain using the second triangular inequality

$$\mathbb{P}(\neg\Omega_k^*|\Omega_{k-1}^*) \leq \mathbb{P}(\Omega_{k-1}^*)^{-1}\mathbb{P}(||\gamma_k(x)| - apn| \geq (\delta_k - \delta_{k-1})(pn)^k|\Omega_k) \tag{2.76}$$

$$\leq (1 - 2(k-1)n^{-\frac{K^2}{9}})^{-1}\mathbb{P}(||\Gamma_k(x)| - apn| \geq 2(\alpha_k + \beta_k + \gamma_k)(pn)^k|\Omega_k) \tag{2.77}$$

19

The last inequality being obtained from the hypothesis of induction and using $(1+x) \leq \exp(x)$. Now using the fact that $apn \leq \frac{3}{2}(pn)^k$ we have

$$\leq 2\mathbb{P}(||\Gamma_k(x)| - apn| \geq (\alpha_k + \beta_k + \gamma_k)apn|\Omega_k) \tag{2.78}$$

$$\leq 2n^{-\frac{K^2}{9}} \tag{2.79}$$

If we combine the last inequality that is obtained applying the previous lemma with (2.74) then we have the result. $\qquad\square$

Definition 2.5.2. For x and y two vertices of G, we say that x and y are remote if $y \notin N_d(x)$.

Proof. From (2.5.3) we can obtain the following equation.

$$\mathbb{P}(|N_{d-1}(x)| < \frac{5}{6}(pn)^{d-1}) < n^{-4} \tag{2.80}$$

Now we estimate the probability that a vertex y is joined to no vertex in a set W with $|W| \geq \frac{5}{6}(pn)^{d-1}$

$$(1-p)^{|W|} \leq \exp(-|W|) \leq \exp(-p|W|) = \exp(-\frac{5}{6}\log(\frac{n^2}{c})) = c^{\frac{5}{6}}n^{-\frac{5}{3}} \tag{2.81}$$

Hence, if x, y, z are vertices we have

$$\mathbb{P}(x \text{ is remote from } y \text{ and } z \text{ }) \tag{2.82}$$

$$\leq \mathbb{P}(|N_{d-1}(x)| \leq \frac{5}{6}(pn)^{d-1}) \tag{2.83}$$

$$+ \mathbb{P}(\{y,z\} \cap N_{d-1}(x) = \emptyset| \ |N_{d-1}(x)| \geq \frac{5}{6}(pn)^{d-1}) \tag{2.84}$$

$$\leq \mathbb{P}(|N_{d-1}(x)| \leq \frac{5}{6}(pn)^{d-1}) \tag{2.85}$$

$$+ (\mathbb{P}(y \text{ is not joined to } W = N_{d-1}(x)|W \geq \frac{5}{6}(pn)^{d-1}))^2 \tag{2.86}$$

$$\leq n^{-4} + c^{\frac{5}{3}}n^{-\frac{10}{3}} \tag{2.87}$$

$$\leq n^{-3}n^{-\frac{1}{4}} \tag{2.88}$$

So, we obtain

$$\mathbb{P}(\mathcal{G}_{n,p} \text{ contains two remote vertices pairs sharing a vertex }) \tag{2.89}$$

$$\leq \sum_x \sum_{(y,z)} \mathbb{P}(x \text{ is remote from } y \text{ and } z \text{ }) \tag{2.90}$$

$$\leq \sum_x \binom{n}{2} n^{-3-\frac{1}{4}} = n \binom{n}{2} n^{-3-\frac{1}{4}} \tag{2.91}$$

$$\leq n^{-\frac{1}{4}} \tag{2.92}$$

The following lemma can be obtained quite easily simply by construction.

20

Lemma 2.5.4. From r disjoint pair of vertices, there are 2^r r-tuples of vertices meeting each pair.

TODO : CHECK FOLLOWING LEMMA

Lemma 2.5.5. The r-th factorial moment of X (X being the number of remote pairs of vertices) is within $o(1)$ of the expected number of ordered r-tuples of disjoint remote pairs.

If we denote by $\mathbb{F}_r$ the probability that a fixed r-tuple consists of vertices remote from each other. Then,

$$\mathbb{E}_r(X) = \frac{n!}{(n-r)!} 2\mathbb{F}_r(1 + o(1)) + o(1) \tag{2.93}$$

In order to shorten the proof, we admit the following lemma that requires a quite long and technical proof. The proof can be found in [Bol81].

Lemma 2.5.6. With probability at least $1 - n^{-K}$

$$(1 - n^{-K})Q_r \le \mathbb{F}_r \le (1 - n^{-K})Q_r + n^{-K} \tag{2.94}$$

With $Q_r = (\frac{c}{n})^r(1 + o(1))$

$$\mathbb{F}_r = (\frac{c}{n})^r(1 + o(1)) \tag{2.95}$$

And we can obtain the asymptotical estimate

$$\mathbb{E}_r(X) = n^r 2^{-r}(\frac{c}{n})^r(1 + o(1)) + o(1) \tag{2.96}$$

Using the following theorem

Theorem 2.5.7. [11]

$$\lim_{n\to\infty} \mathbb{E}_r(X_n) - \lambda^r = 0 \tag{2.97}$$

$$\implies X \longrightarrow_d \mathcal{P}_\lambda \tag{2.98}$$

Then we have that X converges in distribution to Poisson law of parameter $\frac{c}{2}$.
We can then obtain

$$\mathbb{P}(X = 0) = \mathbb{P}(\mathrm{diam}(G) \le d) \longrightarrow e^{-\frac{c}{2}} \tag{2.99}$$

that proves the first part of the theorem. We can also prove the case with $d = 1$ that gives

$$\mathbb{P}(\mathrm{diam}(\mathbb{G}_{n,p}) \le 1) = \mathbb{P}(\mathbb{G}_{n,p} = K_n) = p^{\binom{n}{2}} \longrightarrow 0 \tag{2.100}$$

---

[11]See (A.1.3) for a proof

And to finish the proof we need to consider the two following cases

$$p^{d-1}n^{d-2} = \log(\frac{n^2}{c_1(n)}) \tag{2.101}$$

And in that case we have $c_1 \longrightarrow \infty$.

$$\mathbb{P}(\mathrm{diam}(\mathbb{G}_{n,p}) \le d+1) = e^{-\frac{c_1}{2}} \longrightarrow 0 \tag{2.102}$$

If we now consider

$$p^{d+1}n^d = \log(\frac{n^2}{c_2(n)}) \tag{2.103}$$

we obtain $c_2 \longrightarrow 0$ which gives

$$\mathbb{P}(\mathrm{diam}(\mathbb{G}_{n,p}) \le d+1) = e^{-\frac{c_2}{2}} \longrightarrow 1 \tag{2.104}$$

Finally, if we combine the last result with (2.99) it is clear that the proof if complete. $\qquad\square$

To end this section we present two theorems that are much easier to prove with a similar flavor to the previous one.

Theorem 2.5.8. Suppose

$$i) \quad p^2 n - 2\log(n) \longrightarrow \infty \tag{2.105}$$

$$ii) \quad n^2(1-p) \longrightarrow \infty \tag{2.106}$$

Then almost every graph in $\mathcal{G}_p$ has diameter 2.

Proof.

$$\mathbb{P}(\mathrm{dist}(x,y) > 2) = (1-p^2)^{n-2} \tag{2.107}$$

as it is the probability that two vertices x and y are not connected by a path of length two.
Then using the the first moment method.

$$\mathbb{E}(\text{ pairs not connected by a path of length 2}) \tag{2.108}$$

$$= \binom{n}{2}(1-p^2)^{n-2} \ge \mathbb{P}(\mathrm{diam}(\mathbb{G}_{n,p}) > 2) \tag{2.109}$$

As from $i)$ the LHS converges to zero, we are almost sure that no graph is of diameter more than 2 and then we just have to prove that there is almost no graph of diameter 1. A graph of diameter 1 being a complete graph we have

$$\mathbb{P}(\mathrm{diam}(\mathbb{G}_{n,p}) = 1) = \mathbb{P}(\mathbb{G}_{n,p} = K_n) = p^{\binom{n}{2}} \longrightarrow 0 \tag{2.110}$$

$$\text{iff } \log p^{\binom{n}{2}} \longrightarrow -\infty \tag{2.111}$$

$$\text{iff } n^2 \log p \longrightarrow -\infty \tag{2.112}$$

$$\text{iff } n^2(1-p) \longrightarrow +\infty \tag{2.113}$$

$$\square$$

And simply if we remember the previous theorem (2.3.2) linking $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,M}$ we obtain the following corollary

Corollary 2.5.8.1. If $M = M(n) < \binom{n}{2}$ satisfies

$$\frac{2M^2}{n^3} - \log(n) \longrightarrow \infty \tag{2.114}$$

Then almost every graph in $\mathcal{G}_{n,M}$ is of diameter 2.

Proof. The first condition makes that we are not studying a complete graph then the diameter is not 1. And we have the proof simply using (2.3.2) with $pN = M$ and combining it with the previous theorem (2.5.8) we have the result. $\qquad \square$

Now that we know that the diameter is spread on two values we are interested in restricting it on only one value. We may observe that theorem (2.5.8) gives us enough information to do so. Indeed, we see that if we forced $c \to 0$ then the diameter would be d. This would translate in (2.42) as

$$p^d n^{d-1} - 2\log(n) \longrightarrow \infty \tag{2.115}$$

But this restriction is not sufficient because the values of the diameter might be smaller than d then we will need to make sure that the probability that the diameter is $d - 1$ goes to 0 which we could do forcing $c \to \infty$. In the same fashion as previously this would translate as

$$p^{d-1} n^{d-2} - 2\log(n) \longrightarrow -\infty \tag{2.116}$$

This informal reasoning gives motivation for the following corollary of (2.5.8) that we will now formally prove. [12]

Corollary 2.5.8.2. Suppose $d = d(n) \geq 3$ and $p = p(n)$ satisfy

$$\frac{\log(n)}{d} - 3\log\log n \to \infty \tag{2.117}$$

$$p^d n^{d-1} - 2\log(n) \longrightarrow \infty \quad \text{and} \quad p^{d-1} n^{d-2} - 2\log(n) \longrightarrow -\infty \tag{2.118}$$

Then almost every $\mathbb{G}_{n,p}$ has diameter d.

Proof. Let K_1 and K_2 be positive constants and define $0 < p_1 < p_2 < 1$ by

$$p_1^{d+1} n^d = \log\left(\frac{n^2}{K_1}\right) \quad \text{and} \quad p_2^d n^{d-1} = \log\left(\frac{n^2}{K_2}\right) \tag{2.119}$$

Then

$$\limsup_{n \to \infty} \mathbb{P}(\text{diam}\mathbb{G}_{n,p} \geq d+1) \leq \limsup_{n \to \infty} \mathbb{P}(\text{diam}\mathbb{G}_{n,p_1} \geq d+1) = 1 - e^{-\frac{K_1}{2}} \tag{2.120}$$

---

[12] In fact it is a immediate consequence of the last part of the proof of (2.5.8) but we will prove it only assuming the theorem itself.

$$\limsup_{n\to\infty} \mathbb{P}(\operatorname{diam}\mathbb{G}_{n,p} \le d-1) \le \limsup_{n\to\infty} \mathbb{P}(\operatorname{diam}\mathbb{G}_{n,p_2} \ge d-1) = e^{-\frac{K_2}{2}} \quad (2.121)$$

As we can choose an arbitrary small K_1 and an arbitrarily large K_2 the assertion follows. $\qquad\square$

Now, going back to the degree-diameter problem defined in (2.37), which is of finding the largest graph with fixed diameter and maximal degree. As we already have the Moore upper bound which is best possible, we are interested in finding a lower bound, and maybe hope that it would be close from Moore's bound in order to obtain precise estimate of $n(D, \Delta)$. To obtain such a bound here we will make use of the probabilistic method by sampling on an appropriate graph space (with known diameter and max degree), we will show that almost surely a specified propery doesn't appear (having less that a specified number of vertices).

In other words, by choosing an appropriate function p of n we can obtain some estimates which are almost surely true on D and Δ. Thus n would almost surely be a lower bound of the degree-diameter problem.
In order to have a relation between n, p and Δ we admit the following result from [Bol01].

Theorem 2.5.9. If $\frac{pn}{\log n} \to \infty$, then almost every graph in $\mathcal{G}_{n,p}$ satisfies

$$\Delta(\mathbb{G}_{n,p} = (1 + o(1))pn \qquad (2.122)$$

We can obtain the following result :

Theorem 2.5.10. Suppose $0 < \epsilon < 1$ and the sequence $(D_k), (\Delta_k)$ are such that

$$D_k^4 \le \Delta_k \quad \text{and} \quad D_k \to \infty \qquad (2.123)$$

Then if k is sufficiently large,

$$n(D_k, \Delta_k) \ge \frac{((1-\epsilon)\Delta_k)^{D_k}}{2D_k \log \Delta_k} \qquad (2.124)$$

Proof. As discussed before, to obtain such a proof we need to find the good function p of n, luckily we may also choose to define p on specific values of n in order for the following argument to be true. So, we set

$$n_k = \lceil \frac{((1-\epsilon)\Delta_k)^{D_k}}{2D_k \log \Delta_k} \rceil \qquad (2.125)$$

$$p_k = n_k^{\frac{1}{D_k}-1}(2\log(n_k) + \log\log(n_k))^{1/D_k} \qquad (2.126)$$

From the previous corollary (2.5.8.2), almost every graph in $\mathcal{G}_{n_k,p_k}$ has diameter (at most) D_k. From (2.5.9), almost every graph in $\mathcal{G}_{n_k,p_k}$ has maximal degree

about $p_k n_k$.

$$\Delta(\mathbb{G}_{n_k,p_k}) = (1+o(1))n_k p_k \tag{2.127}$$

$$= (1+o(1))n_k^{1/D_k}(2\log n_k + \log\log n_k)^{1/D_k} \tag{2.128}$$

$$= (1+o(1))\lceil \frac{(1-\epsilon)\Delta_k}{(2D_k\log(\Delta_k))^{1/D_k}}\rceil(2\log n_k + \log\log n_k)^{1/D_k} \tag{2.129}$$

$$\leq \frac{(1+o(1))(1-\epsilon)\Delta_k}{(2D_k\log\Delta_k)^{1/D_k}}(2D_k\log((1-\epsilon)\Delta_k))^{1/D_k} \tag{2.130}$$

$$\leq (1+o(1))(1-\epsilon)\Delta_k \leq (1-\epsilon)\Delta_k \tag{2.131}$$

$$< \Delta_k \tag{2.132}$$

Which gives,

$$n(D_k,\Delta_k) \geq n_k \geq \frac{((1-\epsilon)\Delta_k)^{D_k}}{2D_k\log\Delta_k} \tag{2.133}$$

$\square$

As a strengthening of this result, Bollobás [Bol04] conjectured that for each $\epsilon > 0$, it should be the case that

$$n(D,\Delta) > (1-\epsilon)\Delta^D$$

if Δ and D are sufficiently large.

A quite recent survey on the degree-diameter problem can be found in [MS13].

Chapter 3

Branching processes on random graphs

3.1 Galton-Watson trees and Karp's exploration process

A branching process is the simplest model that can be used to describe the evolution of a population over time. Typically in a branching process we start with one individual, consider it will create a number of individuals through his lifetime. The distribution of this number is called the offspring distribution and we denote it by $\{p_i\}_0^\infty$ such as

$$p_i = \mathbb{P}(\text{ having i child }) \tag{3.1}$$

When we say that a branching process is a "something" branching process, that means that the offspring distribution follows the "something" law (typically a Poisson branching process). And we also denote by Z_n the number of individuals in the n-th generation. Then if we consider that the offspring distribution doesn't depend on the generation of the individual considered we have

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_{n,i} \quad \text{, with } X = \{X_{n,i}\}_{n,i}, \text{ i.i.d.} \tag{3.2}$$

Observing this distribution, we observe that if for some generation k_0 we have $Z_{k_0} = 0$, then $Z_{k_0+k} = 0$ for any k. We would say that that the population dies out at k_0 and one might be interested to study under which condition a population will die out. It was in fact this question that was studied by Galton and Watson (TODO : CHECK AND ADD HISTORICAL DETAILS) that led to the study of branching processes. Hence, we might refer to these branching processes as Galton-Watson processes or trees (GW).
We define $\mu = \mathbb{E}X_{1,1} < \infty$

Lemma 3.1.1. $M_n = \frac{1}{\mu^n} Z_n$ is a Martingale. [1]

Proof. Let's first show the integrability property.

$$\mathbb{E}Z_{n+1} = \mathbb{E}(X_{n,1} + X_{n,2} + ... + X_{n,Z_n}) \tag{3.3}$$

$$= \sum_{k=0}^{\infty} \mathbb{P}(Z_n = k)\mathbb{E}(X_{n,1} + ... + X_{n,k}) \tag{3.4}$$

$$= \sum_{k=0}^{\infty} \mathbb{P}(Z_n = k)\mu k = \mu\mathbb{E}Z_n \tag{3.5}$$

$$= \mu^{n+1}\mathbb{E}Z_0 = \mu^{n+1} \tag{3.6}$$

Then we have that $\mathbb{E}M_{n+1} = 1$.
Now let's prove the property on the conditional expectation.

$$\mathbb{E}(Z_{n+1}|\mathcal{F}_n) = \mathbb{E}(X_{n,1} + X_{n,2} + ... + X_{n,Z_n}|\mathcal{F}_n) \tag{3.7}$$

$$= \mathbb{E}X_{n,1}\mathbb{E}(Z_n|\mathcal{F}_n) = \mu Z_n \tag{3.8}$$

And we obtain,

$$\mathbb{E}(M_{n+1} \mid \mathcal{F}_n) = \mathbb{E}\left(\frac{Z_{n+1}}{\mu^{n+1}}\right) = \frac{Z_n}{\mu^n} = M_n \tag{3.9}$$

$\square$

We now define the following generating function

$$\phi(s) = \sum_{k=0}^{\infty} p_k s^k \quad , |s| < 1 \tag{3.10}$$

Then if we define the extinction probability ζ as $\zeta = \mathbb{P}(\lim_{n\to\infty} Z_n = 0)$ we have

Theorem 3.1.2. The extinction probability is the fixed point of ϕ.

Proof. For this proof we will rewrite the expression for Z_{n+1} as a sum of independent Galton-Watson process. In order to do so we simply consider that the descendency of the second generation is a sum of Z_1 GW process. For all of the $1 \leq j \leq Z_1$ child, we denote by $Z_n(j)$ as the number of descendants of n-th generation of j.

$$Z_{n+1} = \sum_{j=1}^{Z_1} Z_n(j) \tag{3.11}$$

Each of these $Z_n(j)$ is a GW process independent from the others and for each we can then construct $M_n(j)$ as in (3.1.1). As these are Martingales, using the convergence theorem of Martingales (TODO : ADD IT IN ANNEX) we have

---

[1] We take $\mathcal{F}_n = \sigma(X_{i,j}, i, j \in \mathbb{N})$ as the filtration adapted to M_n

that they converge to some random values that we denote by M and $M(j)$. Now if we divide both sides in (3.11) by μ^{n+1}.

$$\frac{Z_{n+1}}{\mu^{n+1}} = M_{n+1} = \mu^{-1} \sum_{j=1}^{Z_1} M_n(j) \tag{3.12}$$

And by taking limits we have

$$M = \frac{1}{\mu} \sum_{j=1}^{Z_1} M(j) \tag{3.13}$$

Which gives that μM is distributed as $\sum_{j=1}^{Z_1}$ and we can finally obtain

$$\mathbb{P}(M = 0) = \mathbb{P}(\mu M = 0) = \mathbb{P}(\sum_{j=1}^{Z_1}(M_j = 0) \tag{3.14}$$

$$= \sum_{k=1}^{\infty} p_k \mathbb{P}(\sum_{j=1}^{k} M(j) = 0 | Z_1 = k) \tag{3.15}$$

$$= \sum_{k=1}^{\infty} p_k (\mathbb{P}(M = 0))^k \tag{3.16}$$

$$= \phi(\mathbb{P}(M = 0)) \tag{3.17}$$

$\square$

Using this result we can obtain the following result

Theorem 3.1.3. The extinction probability of a Poisson branching process of parameter λ is the solution of $x = e^{-\lambda(1-x)}$.

Proof.

$$\phi(s) = \sum_{k=0}^{\infty} p_k s^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} s^k \tag{3.18}$$

$$= e^{-\lambda} e^{\lambda s} = e^{-\lambda(1-s)} \tag{3.19}$$

$\square$

We can obtain the following theorem

Theorem 3.1.4. If $\mathbb{E}X \leq 1$ then the population dies out almost surely.

Proof. If $\lambda < 1$

$$\phi'(s) < \phi'(1) = \lambda < 1 \tag{3.20}$$

The only fixed point is 0. $\square$

Theorem 3.1.5. If $\mathbb{E}X > 1$ then the population survives with a non-zero probability.

Proof. TODO : ADD PROOF □

We will now define an exploration process of such a branching process. We will use the model and notations from Bollobás and Riordan [BR12], which is in fact an extension of the process introduced by Karp in [Kar90]. In this model we consider a graph with n vertices and the exploration will take n steps. For now we will consider the case where we are exploring a connected component. With this process we think of vertices in three different positions, a vertex can be active, that means the algorithm knows the existence of the vertex and is evaluating it. A vertex can be explored, in that case we can consider that the vertex has been completely evaluated and sorted, in some fashion we can forget about it. Otherwise a vertex can be unseen, meaning that we still have no idea of what it is. So in terms of set, we can consider that at time (step) t we have :

$$A_t : \text{ the set of active vertices at time } t \tag{3.21}$$
$$E_t : \text{ the set of explored vertices at time } t \tag{3.22}$$
$$U_t : \text{ the set of unseen vertices at time } t \tag{3.23}$$

The process starts as follows, at $t = 0$, we place one randomly chosen vertex v_0 in A_0, so $U_0 = V \backslash v_0$ and $A_0 = v_0, E_0 = \emptyset$.
For the following steps, at time $t \geq 1$, the process is as follows, a vertex v_t is picked [2] at random in A_{t-1}. For convenience we will add the variable $\eta_t = |N(v_t) \cap U_{t-1}|$ the number of vertices not yet seen that are neighbours from v_t. And then $A_t = (N(v_t) \cap U_{t-1}) \cup A_{t-1} \backslash \{v_t\}$ Finally, we move v_t to E_t and the process stops when all vertices in the connected component of order n are explored $|E_t| = n$, equivalently $t = n$, equivalently $|A_t| = 0$. [3]
This process is called *Breadth first search* and might be referred to as BFS and we may summarise, for $t \geq 1$ it as :

$$A_t = (N(v_t) \cap U_{t-1}) \cup A_{t-1} \backslash \{v_t\} \tag{3.24}$$
$$U_t = U_{t-1} \backslash (\{v_t\} \cup N(v_t)) \tag{3.25}$$
$$E_t = E_{t-1} \cup \{v_t\} \tag{3.26}$$

Connecting this to a Galton-Watson tree, η_t is the direct progeny of v_t so η_t follows the offspring distribution defined by $X_{1,1}$ and we consider that the population dies out only if the algorithm finishes.

$$\begin{cases} |A_0| & = 1 \\ |A_i| & = |A_{i-1}| + \eta_i - 1 = \eta_1 + ... + \eta_i - (i-1) \end{cases} \tag{3.27}$$

---

[2]We use *picked* in the sense that if an element x is picked in X_t then $x \notin X_{t+1}$.
[3]This process can be very easily extended to explore every vertices simply by picking a vertex not yet explored when the set of active vertices is empty

With all the η_i independent and identically distributed random variables. We can then define T as the instant the population dies out.

$$T := \min\{t : A_t = 0\} \tag{3.28}$$

If $T = \infty$ then we say that the population survives.
We have that $\{v_t\}_t$ is a random walk on the tree under exploration and we also get the following evolution equation

$$|S_t| = |S_{t-1}| + \eta_t - 1 \tag{3.29}$$

When we are considering the Erdős Renyi model as a random graph process, we consider that in this random walk each vertex v has a probability p of turning active. Studying the random walk in that case we observe that the number of vertices to which we can connect the vertex v is

$$|U_t| = n - |E_{t-1}| - |A_{t-1}| = n - (t-1) - |A_{t-1}| \tag{3.30}$$

So we have [4]

$$\eta_t \sim \mathrm{Bin}(n - (t-1) - |A_{t-1}|, p) \tag{3.31}$$

(TODO : ADD A VISUAL EXPLANATION OF THE BFS) and we observe comparing it to (3.27) that our random values η_i are no longer independently distributed. However we notice that if A_{t-1} is "small enough" and n "large enough" the r.v. are almost independently distributed. We also denote that the random walk we defined explores only a connected component so intuitively we want to say that if the connected components are small enough and sparse enough, then, they follow a Galton Watson process, hence our random graph would be made of Galton Watson trees. This will be the point of this chapter, studying links between random graphs and branching process, so we will consider that our probability p has to be small enough such that there is not a single connected component. From (2.2.2) we will consider that $p = \frac{\lambda}{n}$.
As the Poisson law is more convenient to work with than the binomial, and also being it's limit in distribution, let's observe through a few theorems the connecting between Poisson and binomial branching process.

Theorem 3.1.6. For a branching process with a binomial offspring distribution with parameter n and p and a branching process with a Poisson offspring distribution with parameter $\lambda = np$

$$\mathbb{P}_{n,p}(T \geq k) = \mathbb{P}^*_\lambda(T^* \geq k) + e_k(n) \quad, \forall k \geq 1 \tag{3.32}$$

With T (resp. T^*) the size of the binomial (resp. Poisson) resulting branching process. And

$$|e_k(n)| \leq \frac{2\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}^*_\lambda(T^* \geq s) \leq \frac{2\lambda^2 k}{n} \tag{3.33}$$

---

[4]This is the conditional law on the number of neighbour of a vertex knowing $|A_{t-1}|$

Proof. The proof makes use of couplings, see Annex (TODO : WRITE ON COUPLINGS IN THE ANNEX), they are a way to define distinct random values on a common probability space giving information on the inter-twinning of both.

$$X_i \sim \text{Bin}(n, \frac{\lambda}{n}) \quad , \quad X_i^* \sim \text{Poi}(\lambda) \tag{3.34}$$

If we denote by $\mathbb{P}$ the joint probability distribution of X_i and X_i^*n then according to the theorem (TODO : ADD THE THEOREM IN ANNEX + PROVE IT) then

$$\mathbb{P}(X_i \neq X_i^*) \leq \frac{\lambda^2}{n} \tag{3.35}$$

Also,

$$\mathbb{P}_{n,p}(t \leq k) = \mathbb{P}(T \geq k, T^* \geq k) + \mathbb{P}(T \geq k, T^* < k) \tag{3.36}$$
$$\mathbb{P}_{\lambda}^*(t \leq k) = \mathbb{P}(T \geq k, T^* \geq k) + \mathbb{P}(T < k, T^* \geq k) \tag{3.37}$$

Which gives,

$$|\mathbb{P}_{n,p}(T \geq k) - \mathbb{P}_{\lambda}^*(T^* \geq k)| \tag{3.38}$$
$$\leq \mathbb{P}(T \geq k, T^* < k) + \mathbb{P}(T < k, T^* \geq k) \tag{3.39}$$

The following part, until the end of the proof, is valid if we exchange T by T^* and X by X^*.
By construction of T, the event $\{T \geq k\}$ is only defined by the events $X_1, ..., X_{k-1}$
Then we have $T \geq k$ and $T^* < k$ if there exists some s such as $X_s \neq X_s^*$. Hence,

$$\mathbb{P}(T \geq k, T^* < k) \leq \sum_{s=1}^{k-1} \mathbb{P}(T \geq K, X_i \neq X_i^*, \forall i \leq s-1, X_s \neq X_s^*) \tag{3.40}$$

If we are in the event,$T \geq K, X_i \neq X_i^*, \forall i \leq s-1$ then

$$X_1^* + ... + X_i^* \geq i, \forall i \leq s-1 \tag{3.41}$$

In particular,

$$X_1^* + ... + X_s^* = T^* - 1 \geq s - 1 \tag{3.42}$$

Then the event $\{T^* \geq s\}$ depends only on the $X_i^*, i \leq s-1$ thus it is independent of the event $X_s \neq X_s^*$. Combining these elements we obtain,

$$\mathbb{P}(T \geq k, T^* < k) \leq \sum_{s=1}^{k-1} \mathbb{P}(T^* \geq s, X_s \neq X_s^*) \tag{3.43}$$

$$\leq \sum_{s=1}^{k-1} \mathbb{P}(T^* \geq s)(X_s \neq X_s^*) \tag{3.44}$$

$$\leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}(T^* \geq s) \tag{3.45}$$

31

The last inequality being obtained by the theorem on couplings (TODO : ADD REFERENCE). Using the remark on the fact that this portion of the proof is valid for both the binomial and the Poisson case we obtain the following inequality which finishes the proof.

$$e_k(n) = |\mathbb{P}_{n,p}(T \geq k) - \mathbb{P}_\lambda^*(T^* \geq k)| \leq \frac{2\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}(T^* \geq s) \qquad (3.46)$$

$\square$

The last theorem gives us some kind of "point wise" convergence between Poisson and binomial branching for trees to be larger than some fixed constant. Now we want to investigate the typical size of a connected component in $\mathcal{G}_{n,p}$. We denote the connected comported of a vertex v by $\mathcal{C}(v)$ and as a typical component we take $\mathcal{C}(1)$.

Theorem 3.1.7. For all $k \geq 1$,

$$\mathbb{P}_{n,p}(|\mathcal{C}(1)| \geq k) \leq \mathbb{P}_{n,p}(T \geq k) \qquad (3.47)$$

Proof. We consider $|U_i| = n - i - |A_i|$ the number of vertices at step $t = i$ and the random variable X_i that is the offspring distribution on vertices that make the branching process stay a tree. And Y_i the converse of X_i that is, the offspring distribution for elements that have already appeared in our branching process. We then have

$$X_i \sim \text{Bin}(|U_{i-1}|, p) \quad , \quad Y_i \sim \text{Bin}(n - |U_{i-1}|, p) \qquad (3.48)$$

And we construct the random variable $X_i^\geq = X_i + Y_i$ Then we have that $X_i \geq \text{Bin}(n, p)$ if X_i and Y_i are independent, so it is true if we consider $X_i \geq$ conditionally on $\{X_j\}_{j=1}^{i-1}$. Which gives that they are i.i.d. variables.
Also, as $Y_i \geq 0$, then $X_j^\geq \geq X_i$. And if we denote by

$$A_i^\geq = X_1 + ... + X_i^\geq - (i - 1) \qquad (3.49)$$

Then,

$$\mathbb{P}_{n,p}(|\mathcal{C}(1)| \geq k) = \mathbb{P}(|A_t| > 0, \forall t \leq k-1) \leq \mathbb{P}(|A_t^\geq| > 0, \forall t \leq k-1) = \mathbb{P}(T \geq k) \qquad (3.50)$$

$\square$

Another similar theorem (TODO : PROVE IT CORRECTLY ??) is the following one that gives a lower bound on the size of a typical connected component although it is less useful than the previous because it is less fit for asymptotic evaluations.

Theorem 3.1.8.
$$\mathbb{P}_{n,p}(|\mathcal{C}(1)| \geq k) \geq \mathbb{P}_{n-k,p}(T \geq k) \qquad (3.51)$$
Here T is the total progeny of a binomial branching process with parameter $n - k$ and p.

Here is a sketch of the proof, the following changes shall just be plugged in the previous proof and the result will appear.

Proof. The proof is very similar to the previous one but the following random variables are used

$$Y_i \sim \text{Bin}(N_{i-1} - (n - k), p) \tag{3.52}$$

then

$$X_i = X_i^{\le} + Y_i \tag{3.53}$$

$\square$

Before finishing this section, here is a theorem that gives the probability law of $|A_t|$ in a random graph branching process.

Theorem 3.1.9.

$$|A_t| + (t - 1) \sim \text{Bin}(n - 1, 1 - (1 - p)^t) \tag{3.54}$$

Proof. Let's first observe by symmetry that

$$X \sim \text{Bin}(m, p) \iff Y = m - X \sim \text{Bin}(m, 1 - p) \tag{3.55}$$

So to prove the theorem we will prove the equivalent statement

$$n - t - |A_t| = |U_t| \sim \text{Bin}(n - 1, (1 - p)^t) \tag{3.56}$$

Indeed,

$$|U_t| = n - t - |A_t| = n - t - |A_{t-1}| - |\eta_t| + 1 \tag{3.57}$$
$$= n - (t - 1) - A_{t-1} - \eta_t \tag{3.58}$$
$$= n - (t - 1) - A_{t-1} - \text{Bin}(n - (t - 1) - |A_{t-1}|, p) \tag{3.59}$$
$$= |U_{t-1}| - \text{Bin}(|U_{t-1}|, p) = \text{Bin}(|U_{t-1}|, 1 - p) \tag{3.60}$$

Induction on the last result gives the expected result. $\square$

3.2 The subcritical case

Now we will apply the results from the previous section in order to prove the following theorem.

Theorem 3.2.1. If $\lambda < 1$
Then

$$\frac{|\mathcal{C}_{\max}|}{\log(n)} \longrightarrow_{\mathbb{P}} I_{\lambda}^{-1} \tag{3.61}$$

Theorem 3.2.2. $\mathbb{P}_{\lambda}(|\mathcal{C}(1)| > t) \le e^{-I_{\lambda} t}$

In the previous theorem, I_λ stands for the large deviation rate function and is defined as follows.

$$I_\lambda = \lambda - 1 - \log(\lambda) \tag{3.62}$$

It is interesting to note that I_λ is positive if $\lambda \neq 0$

Proof of (3.2.2). This proof uses the fact that $|A_t| = 0$ means that the whole connected component has been explored after t steps, so the connected component is of size less than t. Using (3.1.9) we obtain that $|A_t| \sim \mathrm{Bin}(n-1, 1-(1-p)^t) - (t-1)$, so

$$\mathbb{P}_\lambda(|\mathcal{C}(1)| > t) \leq \mathbb{P}(|A_t| > 0) \leq \mathbb{P}(\mathrm{Bin}(n-1, 1-(1-p)^t) \geq t) \tag{3.63}$$

Using Bernoulli's inequality (A.1.4) $1 - (1-p)^t \leq tp$ and observing that for all s positive the following is true

$$\mathbb{P}(\mathrm{Bin}(n-1, 1-(1-p)^t) \geq t) \leq \mathbb{P}(e^{s\mathrm{Bin}(n-1,tp)} \geq e^{st}) \tag{3.64}$$

Then we can apply Markov inequality which gives, $\forall s \geq 0$

$$\mathbb{P}_\lambda(|\mathcal{C}(1)| > t) \leq e^{-st}\mathbb{E}(e^{s\mathrm{Bin}(n,\frac{t\lambda}{n})}) \tag{3.65}$$

Replacing the moment generating function of the binomial with it's value (A.1.1), we obtain

$$\mathbb{P}_\lambda(|\mathcal{C}(1)| > t) \leq e^{-st}(1 - \frac{t\lambda}{n} + e^s\frac{t\lambda}{n})^n \leq e^{-t(s-\lambda e^s + \lambda)} \tag{3.66}$$

Using $s = \log(1/\lambda)$, which minimises the bound [5], we obtain

$$\mathbb{P}_\lambda(|\mathcal{C}(1)| > t) \leq e^{-I_\lambda t} \tag{3.67}$$

$\square$

Now using this result we will obtain a logarithmic bound on the largest connected component.
For this we will use the random variable

$$Z_{\geq k} = \sum_{v \in V} \mathbb{1}_{|\mathcal{C}(v)| \geq k} \tag{3.68}$$

Observing that $Z_{\geq k}$ is equal to 0 if k is larger that the size of the greatest connected component [6] and we denote it by $\mathcal{C}_{\max}$. Hence we have

$$|\mathcal{C}_{\max}| = \max\{k : Z_{\geq k} \geq k\} \tag{3.69}$$

And we obtain

$$\mathbb{E}_\lambda(Z_{\geq k}) = n\mathbb{P}_\lambda(|\mathcal{C}(1)| \geq k) \tag{3.70}$$

Applying the theorem above (3.2.2) we immediately have the following result.

---

[5] Observe that as s must be positive, λ must be smaller than 1 for the argument to be true.

[6] we assume uniqueness of the greatest connected component as asymptotically it is almost surely unique

Lemma 3.2.3. $\mathbb{P}_\lambda(|\mathcal{C}_{\max}| > a\log n) = \mathcal{O}(n^{-\delta}) \quad, a > I_\lambda^{-1}, \quad \delta > 0$

We will now prove the next lemma that is similar to the previous one but gives an upper bound on the greatest connected component instead (in the sub-critical regime).

Lemma 3.2.4. $\mathbb{P}_\lambda(|\mathcal{C}_{\max}| < a\log(n)) = \mathcal{O}(n^{-\delta}) \quad, a \leq I_\lambda^{-1}, \quad \delta > 0$

Proof. This prove will be a little bit more technical as it uses the second moment methods. First of all we will need an estimate of the variance on $Z_\geq$, for this purpose we will use the following function

$$\chi_k(\lambda) = \mathbb{E}_\lambda(|\mathcal{C}(v)|\mathbb{1}_{\{|\mathcal{C}(v)|\geq k\}}) \tag{3.71}$$

Lemma 3.2.5. $\mathbb{V}_\lambda(Z_{\geq k}) \leq n\chi_k(\lambda)$

Proof. By definition of the variance

$$\mathbb{V}_\lambda(Z_{\geq k}) = \sum_{i,j\in V} \left(\mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, |\mathcal{C}(j)| \geq k) - \mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k)\mathbb{P}_\lambda(|\mathcal{C}(j)| \geq k)\right)$$
$$\tag{3.72}$$

And we can split those probabilities as components form an obvious partition of the vertex set as follow

$$\mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, |\mathcal{C}(j)| \geq k) = (\mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, i \leftrightarrow j) + (\mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, |\mathcal{C}(j)| \geq k, i \nleftrightarrow j) \tag{3.73}$$

Furthermore,

$$\mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, |\mathcal{C}(j)| \geq k) = \sum_{l=k}^n \mathbb{P}_\lambda(|\mathcal{C}(i)| = l, |\mathcal{C}(j)| \geq k) \tag{3.74}$$

$$= \sum_{l=k}^n \mathbb{P}_\lambda(|\mathcal{C}(i)| = l)\mathbb{P}(|\mathcal{C}(j)| \geq k||\mathcal{C}(i)| = l) \tag{3.75}$$

And simply observing that in $\mathcal{G}_{n,p}$, the event $\{|\mathcal{C}(j)| \geq k\}$ naturally increases with n, thus decreases if l vertices are removed, we have in the event where i and j are not connected.

$$\mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, |\mathcal{C}(j)| \geq k| i \nleftrightarrow j) \leq \sum_{l=k}^n \mathbb{P}_\lambda(|\mathcal{C}(i)| = l)\mathbb{P}(|\mathcal{C}(j)| \geq k) \tag{3.76}$$

$$\leq \mathbb{P}(|\mathcal{C}(j)| \geq k)\sum_{l=k}^n \mathbb{P}_\lambda(|\mathcal{C}(i)| = l) \tag{3.77}$$

$$\leq \mathbb{P}(|\mathcal{C}(j)| \geq k)\mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k) \tag{3.78}$$

And in the case where i and j are part of the same connected component then

$$\mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, |\mathcal{C}(j)| \geq k| i \leftrightarrow j) = \mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k) \tag{3.79}$$

35

Which gives, removing all negative terms

$$\mathbb{V}(Z_{\geq k}) \leq \sum_{i,j \in V} \mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, i \leftrightarrow j) \tag{3.80}$$

And using properties of the expectation (linearity and behaviour in regard of indicator function) we have

$$\mathbb{V}(Z_{\geq k}) \leq \sum_{i \in V} \sum_{j \in V} \mathbb{E}_\lambda(\mathbb{1}_{|\mathcal{C}(i)| \geq k} \mathbb{1}_{j \in \mathcal{C}(i)}) \tag{3.81}$$

$$\leq \sum_{i \in V} \mathbb{E}_\lambda(\mathbb{1}_{|\mathcal{C}(i)| \geq k} \sum_{j \in V} \mathbb{1}_{j \in \mathcal{C}(i)}) \tag{3.82}$$

$$\leq \sum_{i \in V} \mathbb{E}_\lambda(\mathbb{1}_{|\mathcal{C}(i)| \geq k} |\mathcal{C}(i)|) = n\chi_k(\lambda) \tag{3.83}$$

$$\square$$

As we want to prove that $\mathbb{P}_\lambda(Z_{\geq k_n} = 0)$ goes to 0 for n sufficiently large using Bienaymé-Tchebychev inequality, we will give an upper bound on the variance and a lower bound on the expectation of $Z_{\geq k_n}$.

$$\chi_{k_n}(\lambda) = \sum_{t=k_n}^{n} \mathbb{P}_\lambda(|\mathcal{C}(1)| > t) \tag{3.84}$$

And using $(3.2.2)$[7] we obtain

$$\chi_{k_n}(\lambda) \leq \sum_{t=k_n}^{n} e^{-I_\lambda(t-1)} \leq \frac{e^{-(k_n-1)I_\lambda}}{1 - e^{-I_\lambda}} = \mathcal{O}(n^{-aI_\lambda}) \tag{3.85}$$

And now we need to find a lower bound on the expectation of $Z_{\geq k}$ So we will make use of (3.70) to obtain the formula of the expectation, $(3.1.8)$ to obtain a lower bound and $(3.1.6)$ to convert our lower bound in the Poisson distribution. So we have

$$\mathbb{E}Z_{\geq k} = n\mathbb{P}_\lambda(|\mathcal{C}(1)| \geq k) \geq n\mathbb{P}_{n-k,p}(T \geq k) = n(\mathbb{P}^*_{\lambda_n}(T^* \geq k) + o(1)) \tag{3.86}$$

T and T^* being the total progeny of branching process (Binomial and Poisson), and $\lambda_n = (n-k)p = \frac{n-k}{n}p$ from $(3.1.6)$.
And we can compute the term on the RHS as follows

$$\mathbb{P}^*_{\lambda_n}(T^* \geq k) = \sum_{t=k}^{\infty} \mathbb{P}^*_{\lambda_n}(T^* = k) = \sum_{t=k}^{\infty} \frac{(\lambda_n t)^{t-1}}{t!} e^{-\lambda_n t} \tag{3.87}$$

To simplify the writings we observe $\lambda_n = (1 - o(1))\lambda$ and $I_{\lambda_n} = I_\lambda + o(1)$, then applying Stirling's formula we obtain

$$\frac{(\lambda_n t)^{t-1}}{t!} e^{-\lambda_n t} = \frac{\lambda^{t-1}(1-o(1))^{t-1}t^{t-1}}{t^t} \frac{e^t}{\sqrt{2\pi t}(1+o(1))} e^{-(1-o(1))\lambda t} \tag{3.88}$$

---

[7]The inequality being strict the -1 appears

Simplifying we get

$$\frac{(\lambda_n t)^{t-1}}{t!}e^{-\lambda_n t} = \frac{1}{\lambda t^{\frac{3}{2}}}(\lambda t t^{-\lambda})^t(1+o(1)) = \frac{e^{-I_\lambda t}}{\lambda t^{\frac{3}{2}}}(1+o(1)) \geq \frac{1}{\lambda}e^{-I_\lambda t} \quad (3.89)$$

Now, as the summand is decreasing we can bound it by the integral as follows.

$$n\mathbb{P}^*_{\lambda_n}(T^* \geq k) \geq \frac{n}{\lambda}\int_k^\infty e^{-I_\lambda t}\mathrm{d}t = \frac{e^{-I_\lambda k}}{\lambda I_\lambda} \quad (3.90)$$

With $k_n = a\log(n)$ we have

$$\mathbb{E}Z_{k_n} \geq n\mathbb{P}^*_{\lambda_n}(T^* \geq k_n) \geq n^{-I_\lambda a + 1} \quad (3.91)$$

And finally we have

$$\mathbb{P}(Z_{k_n} = 0) \leq \frac{\mathbb{V}Z_{k_n}}{(\mathbb{E}Z_{k_n})^2} \leq \frac{\mathcal{O}(n^{-I_\lambda a+1})}{n^{-2I_\lambda a+2}\lambda I_\lambda} = \mathcal{O}(n^{I_\lambda a-1}) \quad (3.92)$$

So, if $a < I_\lambda^{-1}$ there is no component larger than $a\log(n)$ asymptotically almost surely. $\qquad\square$

Now, simply combining the previous results (3.2.3) and (3.2.4) wehave our proof of our main theorem in this section (3.2.1). Hence, in $\mathcal{G}_{n,\frac{\lambda}{n}}$ with $\lambda < 1$ the connected components grow at a logarithmic speed. We will see in the next section that this is very different from the case where $\lambda > 1$ in which components grow linearly. This change of speed in the growth rate is called the phase transition.

3.3   The supercritical case : $\lambda > 1$

First we will show that that in the supercritical case there is a component of linear size (in n). In order to show this result we will use the fact that if there exists a path of a linear size, then it is contained in a component of linear size too. Then we will make this result a little bit better with a very similar proof. In fact there exists stronger results that use the same method but make a use of martingales in order to get a really sharp result on the asymptotic size of the greatest connected component, see Bollobás and Riordan [BR12]. (TODO : ADD THE RESULT ? PROVE IT ?)
Here we will consider $p = \frac{1+\epsilon}{n}$. For the beginning we will use the recent approach from Sudakov [KS13] which makes use of an algorithm of graph exploration called the *depth first search* (DFS). (TODO: ADD GRAPHICAL COMPARISON BETWEEN DFS AND BFS) We made use of the breadth first search (BFS) which when exploring a vertex added all the set of neighbours to it's active stack. Here the approach is quite different as instead of adding all of the neighbours, we check for existence of neighbours one by one, and if one is found then the algorithms moves to this new vertex. If no adjacent vertex can

be found then it goes back to the previous vertex.

More formally we will use the same partition of vertex as in the BFS, A_t the *Last-In/First-Out stack* [8] of active vertices, E_t the sorted vertices that we do not have to treat anymore and U_t the vertices that have not yet been added to A_t. The interest of using DFS here is that A_t is by construction always a path. We will describe the behaviour in the case of the exploration of a random graph so we say that we feed our DFS algorithm with $X_N = \{X_i\}_i^N$ a sequence of i.i.d. random values, one for each possible edge. So the algorithm starts at some specified vertex. From there it checks for edges using each time one of the X_i, the number of evaluations is what we refer to as time. If $X_t = 1$, then the new vertex under evaluation is moved from U_t into A_t and the same procedure repeats. In the case that all possible edges from a vertex have been tested and answered negatively, then the vertex is moved to set of explored vertices of corresponding time, so it is moved from A_t to E_t. The algorithm stops when U_t is empty.

In order for the algorithm to be able to explore all components, when A_t is empty, a vertex is selected from U_t.

The proof will make an extensive use of the depth-first search algorithm, at each step of the algorithm, when it is searching for a neighbour it is simply following a Bernoulli random variable of parameter p. So we consider $X = \{X_i\}_i^N$ our sequence of i.i.d. random variables where each X_i follows a Bernoulli of parameter p in order to get an Erdös-Renyi random graph process. So we obtain the following inequality, as in the event $X_i = 1$, then a vertex is simply moved from U_i to A_i. And if there is a sequence of $X_i = 0$ then the vertices might only move from A to E.

$$|A_t \cup E_t| \geq \sum_{i=1}^{t} X_i \tag{3.93}$$

And for the set of active vertices we have

$$|A_t| \leq 1 + \sum_{i=1}^{t} X_i \tag{3.94}$$

From these two inequalities we understand that having knowledge on X will give us knowledge on the behaviour of our depth first search algorithm. And we might make use of this knowledge to obtain information on our connected component.

The following simple lemma will give us information on the behaviour of our binomial. It is the only probabilistic tool that we will use to show our theorem on the growth rate of the giant component.

Lemma 3.3.1. Let $p = \frac{1+\epsilon}{n}$ and $N_0 = \frac{\epsilon n^2}{2}$. Then,

$$|X_{N_0} - N_0 p| \leq n^{\frac{2}{3}} \quad \text{almost surely.} \tag{3.95}$$

---

[8] A LIFO stack is a set in which elements are ordered according to the time they were added in the stack. Only the last element added can be extracted from a LIFO stack. See Knuth TAOCP I (TODO: ADD REF) for a detailed review of stacks, lists, ...

Proof. Let's observe that $\mathbb{E}X_{N_0} = N_0 p = \frac{\epsilon(1+\epsilon)}{2}n$. Simply using Chernoff inequality (A.2.1.2).

$$\mathbb{P}(|X_{N_0} - \mathbb{E}X_{N_0}| > n^{\frac{2}{3}}) \leq 2e^{-\frac{n^{\frac{4}{3}}}{2n}} \tag{3.96}$$

So asymptotically almost surely we have

$$|X_{N_0} - \mathbb{E}X_{N_0}| \leq n^{\frac{2}{3}} \tag{3.97}$$

$\square$

Now we can state and prove the following theorem making use of the previous lemma and of our knowledge of the depth first search algorithm.

Theorem 3.3.2. Let $p = \frac{1+\epsilon}{n}$ and $N_0 = \frac{\epsilon n^2}{5}$. Then, $G \sim \mathcal{G}_{n,p}$ contains a path of length at least $\frac{\epsilon^2 n}{5}$.

Proof. The proof will be done by contradiction.
We consider X_N with parameter $p = \frac{1+\epsilon}{n}$. We claim that if $N_0 = \frac{\epsilon n^2}{2}$ then $|A_{N_0}| \geq \frac{\epsilon^2 n}{5}$.
Let's first show that $E_{N_0} < \frac{n}{3}$.
If it was not the case, elements flowing in E one by one, there would exist a t such that $E_t = \frac{n}{3}$. Also from (3.94) and (3.3.1) we would have

$$|A_t| \leq 1 + \sum_{i=1}^{t} X_i < 1 + n^{\frac{2}{3}} < \frac{n}{3}. \tag{3.98}$$

Then using the fact that the sets used in the depth first search do not intersect we have

$$|U_t| = n - |A_t| - |E_t| \geq \frac{n}{3} \tag{3.99}$$

So, we obtain that the algorithm has tested all the $|E_t||U_t| \geq \frac{n^2}{9}$ possible pairs between the set of explored vertices and not seen vertices. But $\frac{n^2}{9} > \epsilon \frac{n^2}{2} = N_0$ [9] and as we assumed that we are at a step t of the algorithm that is less than N_0 we have a contradiction.
We are then sure from the previous argument that $|E_{N_0}| < \frac{n}{3}$ and we claim $|A_{N_0}| < \frac{\epsilon^2 n}{5}$, then $U_{N_0} \neq \emptyset$. Which means that there are still elements that can be added to the connected component. We are going to use the same arguments as previously.
By lemma (3.3.1), the number of edges (or vertices) added is at least $\frac{\epsilon(1+\epsilon)n}{2} - n^{\frac{2}{3}}$. Which gives that the number of active and explored vertices is at least as follows

$$|A_{N_0} \cup E_{N_0}| \geq \frac{\epsilon(1+\epsilon)n}{2} - n^{\frac{2}{3}} \tag{3.100}$$

---

[9] ϵ is small enough

So $|E_{N_0}| \geq \frac{\epsilon n}{2} + \frac{3\epsilon^2 n}{10} - n^{\frac{2}{3}}$ and that would mean all of the pairs between E_{N_0} and A_{N_0} have been explored. So we obtain the following set of inequalities.

$$N_0 = \frac{\epsilon n^2}{2} \geq |E_{N_0}||A_{N_0}| \geq (\frac{\epsilon n}{2} + \frac{3\epsilon^2 n}{10} - n^{\frac{2}{3}})(n - \frac{\epsilon n}{2} - \frac{\epsilon^2 n}{2} + n^{\frac{2}{3}}) \quad (3.101)$$

$$\geq \frac{\epsilon n^2}{2} + \frac{\epsilon^2 n^2}{20} - o(\epsilon^3)n^2 \quad (3.102)$$

The last inequality being only with the dominating terms (i.e. the n^2) and for $\epsilon < 1$ it is larger than $\frac{\epsilon n^2}{2}$ Which is a contradiction. Then A_{N_0} must be larger or equal than $\frac{\epsilon^2 n}{5}$ and observing that A_{N_0} must be a path we have the result. $\square$

In fact, we can refine this result to the size of the connected component and not simply the size of a walk with the following theorem.

Theorem 3.3.3. Let $p = \frac{1+\epsilon}{n}$ and $N_0 = \frac{\epsilon n^2}{5}$. Then, $G \sim \mathcal{G}_{n,p}$ contains a connected component of size at least $\frac{\epsilon n}{2}$

Proof. (TODO : ADD PROOF) $\square$

3.4   Some words on the critical case

Chapter 4

The configuration model

4.1 Generalized Binomial Graph

We are now interested in a generalization of the (binomial) Erdös-Rényi model which was first considered by Kovalenko in (TODO : ADD REF). In this model, often referred to as *inhomogeneous*, the probability of appearance of the edge (i, j), that we call p_{ij}, is not necessarily the same for all pairs of vertices i, j. From this definition it is very natural to write those edge probabilities in an $n \times n$ matrix, which we denote by

$$\mathbf{P} = [p_{ij}] \qquad (4.1)$$

As previously we consider graphs without loops, so $p_{ii} = 0$ for all $i \in V$, we also consider that the graph is not directed so the probability that i connects to j has to be the same as the probability that j connects to i (i.e. $p_{ij} = p_{ji}$), hence $\mathbf{P}$ is symmetric. In order to shorten the writing, we put $q_{ij} = 1 - p_{ij}$.

We denote the probability space of this generalized model as $\mathcal{G}_{n,\mathbf{P}}$ and as before we denote a random variable in this space as $\mathbb{G}_{n,\mathbf{P}}$.

We now define

$$Q_i = \prod_{j=1}^{n} q_{ij}, \quad \lambda_n = \sum_{i=1}^{n} Q_i \qquad (4.2)$$

and we observe that Q_i is the probability that vertex i is isolated, hence λ_n is the expected number of isolated vertices.

We also need to define

$$R_{ik} = \min_{1 \leq j_1 < j_2 < \ ...<j_k \leq n} q_{ij_1} q_{ij_2} \ldots q_{ij_k} \qquad (4.3)$$

In the following, we suppose that the edge probabilities p_{ij} are chosen in such a way that the following equations are simultaneously satisfied as $n \to \infty$.

$$\max_{1 \leq i \leq n} Q_i \to 0 \tag{4.4}$$

$$\lim_{n \to \infty} \lambda_n = \lambda = \text{constant} \tag{4.5}$$

$$\lim_{n \to \infty} \sum_{k=1}^{n/2} \frac{1}{k!} \left(\sum_{i=1}^{n} \frac{Q_i}{R_{ik}} \right)^k = e^\lambda - 1 \tag{4.6}$$

In this section we are interested in proving the following theorem.

Theorem 4.1.1. Let X_0 denote the number of isolated vertices in $\mathbb{G}_{n,\mathbf{P}}$. If (4.4), (4.5), (4.6) are satisfied, then the number of isolated vertices is asymptotically Poisson distributed with mean λ.
i.e. for $k \in \mathbb{N}$,

$$\lim_{n \to \infty} \mathbb{P}(X_0 = k) = \frac{\lambda^k}{k!} e^{-\lambda} \tag{4.7}$$

Let's first show as a corollary that this theorem allows us to conclude the proof of 2.2.1.

Corollary 4.1.1.1. If $p(n) = \frac{\log n + c}{n}$, then (4.4), (4.5), (4.6) are satisfied, and the number of isolated vertices is asymptotically Poisson distributed with mean e^{-c}.

Proof of 4.1.1.1. We observe that $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,\mathbf{P}}$ when $p_{ij} = p$ for all $i \neq j$. We observe that $Q_i = Q = q^{n-1}$ (and (4.4) is obviously satisfied) and $R_{ik} = R_k = q^k$. From this we get

$$\lambda_n = nQ = nq^{n-1} = n(1-p)^{n-1} \tag{4.8}$$

and we have $\lim \lambda_n = e^{-c} = \lambda$, so (4.5) is also satisfied.
Let's observe that it is enough to show $\sum_{i=1}^{n} \frac{Q_i}{R_{ik}}$ converges uniformly to $\lambda = e^{-c}$ for (4.6) to be satisfied.

$$\sum_{i=1}^{n} \frac{Q_i}{R_{ik}} = \sum_{i=1}^{n} \frac{Q}{R_k} = n\frac{Q}{R_k} = nq^{n-1-k} \tag{4.9}$$

And $\lim_{n \to \infty} nq^{n-1-k} = e^{-c}$ finally proves the corollary. $\square$

Proof. Proof of 4.1.1 Let Y_{ij} a random variable following a Bernoulli of parameter p_{ij}. [1] We denote by Y_i the indicator of the event that vertex i is isolated, i.e. $Y_i = \mathbb{1}_{\sum_{j=1}^{n} Y_{ij} = 0}$. In order to show the convergence of X_0 in distribution to the Poisson random variable we will use the method of factorial moments. So, we want to show that for any natural number k we have

$$\mathbb{E}\left(\sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} X_{i_1} X_{i_2} \ldots X_{i_k} \right) \to_{n \to \infty} \frac{\lambda^k}{k!} \tag{4.10}$$

---

[1] Recall that a Bernoulli of parameter p is 1 with prob. p and 0 with prob. $1 - p$.

Let's observe that the LHS of (4.10) is the sum of $\mathbb{E}(X_{i_1} X_{i_2} \ldots X_{i_k})$ over all $i_1 < i_2 < \ldots i_k$.

$$\mathbb{E}(X_{i_1} X_{i_2} \ldots X_{i_k}) = \prod_{r=1}^{k} \mathbb{P}(X_{i_r} = 1| \ X_{i_1} = \ldots = X_{i_{r-1}} = 1) \tag{4.11}$$

$$= \prod_{r=1}^{k} \frac{\prod_{j=1}^{n} q_{i_r j}}{\prod_{s=1}^{r-1} q_{i_r i_s}} \tag{4.12}$$

From which we get,

$$Q_{i_r} \le \mathbb{P}(X_{i_r} = 1| \ X_{i_1} = \ldots = X_{i_{r-1}} = 1) \le \frac{Q_{i_r}}{R_{i_r,r-1}} \le \frac{Q_{i_r}}{R_{i_r k}} \tag{4.13}$$

Hence,

$$Q_{i_1} \ldots Q_{i_k} \le \mathbb{E}(X_{i_1} \ldots X_{i_k}) \ \le \frac{Q_{i_1}}{R_{i_1 k}} \ldots \frac{Q_{i_k}}{R_{i_k k}} \tag{4.14}$$

TODO : WHY IS THE FOLLOWING INEQ TRUE ?

$$\sum_{1 \le i_1 < \ldots < i_k \le n} Q_{i_1} \ldots Q_{i_k} = \frac{1}{k!} \sum_{1 \le i_1 \ne \ldots \ne i_r \le n} Q_{i_1} \ldots Q_{i_k} \ge \tag{4.15}$$

$$\frac{1}{k!} \sum_{1 \le i_1, \ldots, i_k \le n} Q_{i_1} \ldots Q_{i_k} - \frac{k}{k!} \sum_{i=1}^{n} Q_i^2 (\sum_{1 \le i_1, \ldots, i_{k-2} \le n} Q_{i_1} \ldots Q_{i_{k-2}}) \tag{4.16}$$

$$\ge \frac{\lambda_n^k}{k!} - (\max_i Q_i)\lambda_n^{k-1} \longrightarrow \frac{\lambda^k}{k!} \tag{4.17}$$

Now, by definition of R_{ik}, we have,

$$\sum_{i=1}^{n} \frac{Q_i}{R_{ik}} \ge \sum_{i=1}^{n} Q_i = \lambda_n \tag{4.18}$$

If we suppose that $\limsup_{n \to \infty} \sum_{i=1}^{n} \frac{Q_i}{R_{ik}} > \lambda$, this would imply that $\limsup_{n \to \infty} \sum_{k=1}^{\frac{n}{2}} \frac{1}{k!}(\sum_{i=1}^{n} \frac{Q_i}{R_{ik}})^k > e^{\lambda} - 1$ which contradicts (4.6). It follows that

$$\lim_{n \to \infty} \sum_{i=1}^{n} \frac{Q_i}{R_{ik}} = \lambda \tag{4.19}$$

We obtain,

$$\sum_{1 \le i_1 < \ldots < i_k \le n} Q_{i_1} \ldots Q_{i_k} \le \frac{1}{k!}(\sum_{i=1}^{n} \frac{Q_i}{R_{ik}})^k \longrightarrow \frac{\lambda^k}{k!} \tag{4.20}$$

Combining (4.20) and (4.15) we have $\lim_{n \to \infty} \sum_{1 \le i_1 < \ldots < i_k \le n} Q_{i_1} \ldots Q_{i_k} = \frac{\lambda^k}{k!}$, which is the result we wanted to prove. □

4.2   An arbitrary degree sequence - the Newman-Watts-Strogatz model

Appendix A

Some probabilistic tools

A.1 Common inequalities and simple probabilistic results

This section presents inequalities that are used in this report and proves most of them. (TODO : PROVE THEM)

Theorem A.1.1 (The moment generating function of a binomial)**.** If $X \sim \text{Bin}(n, p)$, then

$$\mathbb{E}(e^{tX}) = (1 - p + e^t p)^n \tag{A.1}$$

Proof.

$$\mathbb{E}(e^{tX}) = \sum_{k=1}^{n} e^{tk} \mathbb{P}(X = k) \tag{A.2}$$

$$= \sum_{k=1}^{n} \binom{n}{k} e^{tk} p^k (1 - p)^k \tag{A.3}$$

$$= \sum_{k=1}^{n} \binom{n}{k} (e^t p)^k (1 - p)^k \tag{A.4}$$

$$= (1 - p + e^t p)^n \tag{A.5}$$

$\square$

Theorem A.1.2 (Convergence of factorial moments implies convergence in distribution)**.** Let X be a random variable with a distribution that is determined by its moments. If $X_1, X_2, \ldots$ are random variables with finite moments such that $\mathbb{E}_k(X_n) \longrightarrow \mathbb{E}_k(X)$ when $n \to \infty$ for every integer $k \geq 1$ then

$$X_n \longrightarrow_d X \tag{A.6}$$

Theorem A.1.3 (Factorial moments of a Poisson). If $X \sim \text{Poi}(\lambda)$ then

$$\mathbb{E}_r(X) = \lambda^r \tag{A.7}$$

Theorem A.1.4 (Bernoulli's inequality).

$$1 - (1-p)^t \leq tp \tag{A.8}$$

Proof. We show the inequality is true for any $k \in \mathbb{N}^*$ by induction. The result follows by convexity.

It is clearly true fir $k = 1$.

Assume the inequality is true at rank k, then,

$$1 - (1-p)^{t+1} = 1 - (1-p)^t(1-p) = 1 - (1-p)^t - p(1-p)^t \tag{A.9}$$

$$\leq p(t - (1-p)^t) \leq p(t + 1 - (1-p)^t) \leq p(1+t) \tag{A.10}$$

which proves the result for any $t > 0$. □

Theorem A.1.5 (Stirling's formula).

$$n! \sim (\frac{n}{e})^n \sqrt{2\pi n} \tag{A.11}$$

Corollary A.1.5.1 (n choose k approximation).

$$\binom{n}{k} \leq \frac{n^k}{k!} \leq (\frac{ne}{k})^k \tag{A.12}$$

Proof.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{A.13}$$

$$\leq \frac{n^n \sqrt{(n)} e^{-n}}{\sqrt{n-k} e^{-(n-k)} (n-k)^{n-k} k!} \leq \frac{n^k}{k!} \tag{A.14}$$

$$\leq \frac{n^k}{\sqrt{2\pi k} k^k e^{-k}} = (\frac{ne}{\sqrt{2\pi k} k})^k \tag{A.15}$$

□

Theorem A.1.6 (Markov's inequality).

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a} \tag{A.16}$$

Proof.

$$\mathbb{E}X = \int_0^\infty x f(x) dx \geq \int_a^\infty x f(x) dx \geq a \int_a^\infty f(x) dx = a\mathbb{P}(X \geq a) \tag{A.17}$$

□

Corollary A.1.6.1 (Chebyshev's inequality).

$$\mathbb{P}(|X - \mathbb{E}X| \geq k\mathbb{V}X) \leq \frac{1}{k^2} \tag{A.18}$$

Proof. Apply (A.1.6) using the random variable $(X - \mathbb{E}X)^2$ □

A.2   Tail inequalities

The following inequalities are designated as Chernoff inequalities.

Theorem A.2.1 (Chernoff bound)**.** If $X \sim \text{Bin}(n, p)$, and $\lambda = np$ then $\forall t \geq 0$

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq e^{-\frac{t^2}{2(\lambda + t/3)}} \tag{A.19}$$

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq e^{-\frac{t^2}{2\lambda}} \tag{A.20}$$

Corollary A.2.1.1. If $X \sim \text{Bin}(n, p)$, and $\lambda = np$ then $\forall t \geq 0$

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq 2e^{-\frac{t^2}{2(\lambda + t/3)}} \tag{A.21}$$

Corollary A.2.1.2. If $X \sim \text{Bin}(n, p)$, and $0 < \epsilon \leq 3/2$ then

$$\mathbb{P}(|X - \mathbb{E}X| \geq \epsilon \mathbb{E}X) \leq 2e^{-\frac{\epsilon^2}{3}\mathbb{E}X} \tag{A.22}$$

Theorem A.2.2 (Binomial tail)**.**

A.3   Markov chains

A.4   Martingales

References

[BT87]   B. Bollobás and A. G. Thomason. "Threshold functions". In: *Combinatorica* 7.1 (Mar. 1, 1987), pp. 35–38. ISSN: 1439-6912. DOI: `10.1007/BF02579198`. URL: `https://doi.org/10.1007/BF02579198`.

[Bol04]   Béla Bollobas. *Extremal Graph Theory*. New York, NY, USA: Dover Publications, Inc., 2004. ISBN: 0486435962.

[Bol81]   Béla Bollobás. "The Diameter of Random Graphs". In: *Transactions of the American Mathematical Society* 267.1 (1981), pp. 41–52. ISSN: 00029947. URL: `http://www.jstor.org/stable/1998567`.

[Bol01]   Béla Bollobás. *Random Graphs*. 2nd ed. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2001. DOI: `10.1017/CBO9780511814068`.

[BR12]   Béla Bollobás and Oliver Riordan. "Asymptotic normality of the size of the giant component in a random hypergraph". In: *Random Structures & Algorithms* 41.4 (2012), pp. 441–450. DOI: `10.1002/rsa.20456`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.20456`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20456`.

[BT85]   Béla Bollobás and Andrew Thomason. "Random Graphs of Small Order". In: *Random Graphs '83*. Ed. by Michał Karoński and Andrzej Ruciński. Vol. 118. North-Holland Mathematics Studies. North-Holland, 1985, pp. 47–97. DOI: `https://doi.org/10.1016/S0304-0208(08)73612-0`. URL: `http://www.sciencedirect.com/science/article/pii/S0304020808736120`.

[BM08]   J.A. Bondy and U.S.R Murty. *Graph Theory*. 1st. Springer Publishing Company, Incorporated, 2008. ISBN: 1846289696.

[Cal+01] Duncan S. Callaway et al. "Are randomly grown graphs really random?" In: 64.4 (Oct. 2001), p. 041902. DOI: `10.1103/PhysRevE.64.041902`. arXiv: `cond-mat/0104546 [cond-mat.stat-mech]`.

[CL01]   Fan Chung and Linyuan Lu. "The Diameter of Sparse Random Graphs". In: *Advances in Applied Mathematics* 26.4 (2001), pp. 257–279. ISSN: 0196-8858. DOI: `https://doi.org/10.1006/aama.2001.0720`. URL: `http://www.sciencedirect.com/science/article/pii/S0196885801907201`.

[ER59]   Paul Erdös and Alfréd Rényi. "On random graphs I". In: *Publ. Math. Debrecen* 6 (1959), pp. 290–297.

[ER60]   Paul Erdös and Alfréd Rényi. "On the evolution of random graphs". In: *Publ. Math. Inst. Hung. Acad. Sci* 5 (1960), pp. 17–61.

[ER61a]   Paul Erdös and Alfréd Rényi. "On the evolution of random graphs". In: *Bull. Inst. Internat. Statist.* 38 (1961), pp. 343–347.

[ER61b]   Paul Erdös and Alfréd Rényi. "On the strength of connectedness of a random graph". In: *Acta. Math. Acad. Sci. Hungar* 12 (1961), pp. 261–267.

[ER63]   Paul Erdös and Alfréd Rényi. "Asymmetric graphs". In: *Acta. Math. Acad. Sci. Hungar* 14 (1963), pp. 295–315.

[Gil59]   E. N. Gilbert. "Random Graphs". In: *Ann. Math. Statist.* 30.4 (Dec. 1959), pp. 1141–1144. DOI: `10.1214/aoms/1177706098`. URL: `https://doi.org/10.1214/aoms/1177706098`.

[JŁR00]   Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random graphs0*. John Wiley and Sons, 2000. DOI: `10.1002/9781118032718`.

[Joy81]   André Joyal. "Une théorie combinatoire des séries formelles". In: *Advances in Mathematics* 42.1 (1981), pp. 1–82. ISSN: 0001-8708. DOI: `https://doi.org/10.1016/0001-8708(81)90052-9`.

[Kar90]   Richard M. Karp. "The transitive closure of a random digraph". In: *Random Structures & Algorithms* 1.1 (1990), pp. 73–93. DOI: `10.1002/rsa.3240010106`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.3240010106`.

[KS13]   Michael Krivelevich and Benny Sudakov. In: *Random Structures & Algorithms* 43.2 (2013), pp. 131–138. DOI: `10.1002/rsa.20470`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20470`.

[LZ12]   Gyu Eun Lee and Doron Zeilberger. "Joyal's proof of Cayley's formula". In: (2012). URL: `http://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimPDF/JoyalCayley.pdf`.

[MS13]   Mirka Miller and Jozef Sirán. "Moore Graphs and Beyond: A survey of the Degree/Diameter Problem". In: 2013. URL: `https://www.combinatorics.org/files/Surveys/ds14/ds14v2-2013.pdf`.

[NSW01]   M. E. J. Newman, S. H. Strogatz, and D. J. Watts. "Random graphs with arbitrary degree distributions and their applications". In: 64, 026118 (Aug. 2001), p. 026118. DOI: `10.1103/PhysRevE.64.026118`. arXiv: `cond-mat/0007235 [cond-mat.stat-mech]`.

[NWS02]   M. E. J. Newman, D. J. Watts, and S. H. Strogatz. "Random graph models of social networks". In: *Proceedings of the National Academy of Sciences* 99.suppl 1 (2002), pp. 2566–2572. ISSN: 0027-8424. DOI: 10.1073/pnas.012582999. eprint: `https://www.pnas.org/content/99/suppl_1/2566.full.pdf`. URL: `https://www.pnas.org/content/99/suppl_1/2566`.

[Spe14]   Joel Spencer. *Asymptopia*. English (US). American Mathematical Society, 2014.

[Sva14]   Janson Svante. "random graphs". In: 2014. URL: `http://www2.math.uu.se/~svante/talks/2014stockholm.pdf`.

[Tru93]   R.J. Trudeau. *Introduction to Graph Theory*. Dover Books on Mathematics. Dover Pub., 1993. ISBN: 9780486678702.

[WS98]   Duncan J. Watts and Steven H. Strogatz. "Collective dynamics of 'small-world' networks". In: *Nature* 393.6684 (June 1998), pp. 440–442. ISSN: 0028-0836. DOI: 10.1038/30918. URL: `http://dx.doi.org/10.1038/30918`.