On random graphs

Leo Davy

March 2019

Contents

Chapter 1

An introduction to random graphs

1.1   Graph theory

Intuitively, a graph is just about dots and lines, any phenomenon that can be represented as dots connected, or not, by lines can be thought of as a graph. Hence it is clear that graph theory, the study of the so called graphs that we will define in the following, will apply to a very wide variety of problems, such as, epidemiology, sociology, internet analysis, electric circuits, road traffic, ... and of course mathematics. Historically graphs first appeared in mathematics from Leonard Euler who gave, again, his name to a formula that would be the basis for the development of topology. S

Formally a graph G will be defined as $G = (V(G), E(G))$, with $V(G)$ designing the vertex set of G and $E(G)$, disjoint from $V(G)$, the set of edges of G. For now we will consider edges only as a pair of elements, without order, contained in $V(G)$. In the following of this report hypergraphs will be studied in which edges are composed of more than two elements.

As an example of a graph we can consider the following graph with $V(G) = \{a, b, c, d, e\}, E(G) = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8\}$ altogether with an incidence function $\psi_G : V(G) \rightarrow E(G)$ such as:

$$\psi_G(e_1) = ab \quad \psi_G(e_2) = ac \quad \psi_G(e_3) = bc \quad \psi_G(e_4) = ad$$
$$\psi_G(e_5) = cd \quad \psi_G(e_6) = cd \quad \psi_G(e_7) = ee \quad \psi_G(e_8) = ae$$

An equivalent way to define a graph would be through the incidence matrix $M_G = (m_{ve})$ zith m_{ve} the number of times the vertice v and the edge e are incident. So m_{ve} can take the values 0 (not incident), 1 or 2 (e is a loop). It is also possible to define a graph in a third way, that is equivalent for the structure but doesn't take in account the labelling of the edges, it is through the adjacency matrix $A_G = (a_{uv})$. This is usually the most useful version as typically a graph will have less vertices than edges, the adjacency matrix will

be much smaller to write (hence to store in a computer) and it usually gives all the information needed to study a graph. It's interesting to note that an adjacency matrix is real and hermitian, hence all of it's eigenvalues are real and the study of it's distribution is a common topic in graph theory.

An interesting property of graphs is the degree of the vertices, so we will denote by $d_G(v)$ the number of edges incident with $v \in V(G)$. 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. From these definition we can obtain the following lemma, with m the number of edges.

Theorem 1.1.1. For any graph finite graph G

$$\sum_{v \in V(G)} d_G(v) = 2m \tag{1.1}$$

Proof. The sum of the elements of each columns in the incidence matrix is equal to two. So the sum of the values in the colums over all the columns is equal to two times the number of columns, so $2m$. As the sum of the columns is equal to the sum of the rows, and the sum of each row is exactly the degree of a vertex, we have the result. $\square$

This theorem will prove useful in the following as it connects the number of edges and the degrees of the vertices. In graph theory a graph can be represented in many different ways and then it can be really non trivial to know if two graphs with different labelling are the same. More formally, for a same graph, we will call the set of permutation of the labellings that doesn't change the structure of the graph, it's automorphism group, denoted $Aut(G)$ and $aut(G)$ it's cardinal. And for the anecdote, finding the problem of showing that two graphs are in the same automorphism group is a NP-hard problem.

A very important kind of graphs are the directed graphs, it's a very intuitive notion, these are simply graphs on which there are arrows on the edges, equivalently, it is like defining the adjacency between two vertices as non symmetric, or as studying graph with non hermitian adjacency matrix

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

1.2   Random graphs

This section will try to give reasons behind the study of random graphs but it's purpose is not to go deep in the details and sophistication of their study. The study of random graphs is a flourishing area of mathematics since it's founding papers have been published by Erdos and Renyi between 1959 and 1963. Since then a lot of work has been done on random graphs, most of the questions on the Erdos-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.

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

Random graphs are interesting subject for pure mathematicians as they create a lot of open problems and offers many links with combinatorics. And for an applied mathematicians, random graphs are an entertaining tool as they may be used to simulate real world phenomenons (most famously in sociology, epidemiology or the study of internet). And accordingly to their level of matching with real life situations, they will be able to show the presence of complexity in the situation studied.

1.3   Cayley's formula

This section will first of all demonstrate an important result that will be used several times in crucial demonstrations in this report. Although it is not a demonstration that is specific to random graphs it may give an insight to the variety of techniques that may be used in the study of random graphs and how elegant are the results (at least quite often). This formula has been demonstrated in many different ways and we will use the demonstration by Joyal that is really elegant and also is a good place to introduce several notions that will be used in the rest of the report.

Theorem 1.3.1. Cayley's formula

$$t_n = n^{n-2} \tag{1.2}$$

with t_n the number of spanning trees on n vertices.

Before beginning the proof some definitions will be needed. A structure (graph or tree) is called spanning on the vertices (resp. edges) if it intersects all vertices (resp. edges). 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 it's edges it becomes disconnected, equivalently, it's a graph in which every two vertices are linked by exactly one path, equivalently, it's a connected and acyclic graph (doesn't contain any cycle).
The trees being a subset of the graphs, it is also possible to define directed trees in which you can follow an edge only in one direction (otherwise it would not be a tree anymore). We also define doubly rooted trees as trees with two special labels "Start" and "End" that can be attached to any vertices of the tree and which canonically maps on each edges the direction such that any vertice can reach the end. And we will call "SEL" the vertices that are in the "Start" to "End" line. 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 would then be sufficient to prove that the number of elements in DRT_n is equal to n^n. So we will base our approach on Joyal's proof and show a bijection between the set of doubly rooted trees on n vertices and the the set of functions on n elements.

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 $v_1 \rightarrow v_2$ 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 cycles. And the whole graph will be a disjoint union of such structures. 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 and now we will construct a graph with the vertice set $D = f(C)$, and the edge set $E = \bigcup_{i=1}^{k-1} f(c_i)f(c_{i+1})$. 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".

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 vertice j, we define $f(j)$ as the first neighbour on the j to end line.

For vertices on the SEL,

$$SEL = \{x_1, x_2, ..., x_k\}, \text{and} SEL_< = \{x_{\sigma_1}, x_{\sigma_2}, ..., x_{\sigma_k}\} \tag{1.3}$$

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 Erdos-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.

Historically the first paper published on random graphs was from Erdos and Renyi in 1959, in which they give the following construction :

Definition 2.1.1. We call a random graph $\mathcal{G}_{n,M}$ having n possible 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 Erdos 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 fair, but the probability of drawing an edge is now p. 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 :

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

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. 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 Erdos and Renyi. The advantage of this model is that it allows one to study properties on the verge of their realisations. For instance, using this model Bollobas proved that a graph is fully connected, when the last connection made is between an isolated vertice and the giant component. But we will study this in the following.

2.2 Connectivity

One of the most fundamental structure of a graph will be it's number of components. Hence, the first question we will try to ask is how often a random graph in the Erdos-Renyi model is connected. It is essential to answer this question as many other questions might not make sense 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 didactic way as it is the first random graphs proof that we will study.

Theorem 2.2.1. Let $p = p(n) = \frac{log(n)+c}{n}$
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.

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

Proof. During this proof we will consider the random value X_k that counts the number of isolated vertices of order 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.

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

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

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

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

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

So it's sufficient that $p = o^{-2}$ 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 other than 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 connected to 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. 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 series on k will converge.

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

$$\leq n^k e^k k^{-2} p^{k-1} e^{-(\log(n)+c)\frac{k(n-k)}{n}} \tag{2.11}$$

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

$$\leq s \tag{2.13}$$

$\square$

2.3 Existence of thresholds

2.4 The stability number

2.5 The diameter

Chapter 3

Branching processes on random graphs

Chapter 4

The configuration model

Appendix A

Some probabilistic tools

A.1   Common inequalities

A.2   Tail inequalities

A.3   Markov chains

A.4   Martingales