Random graph models and their applications

Vallier, Thomas

2007

Acknowledgements

Contents

List of papers

This thesis is based on the following four papers:

A. T. Vallier: Robustness of Preferential Attachment under Deletion of Edges, *Stochastic Models*, **23** 265-276 (2007)

B. T. Turova, T. Vallier: Merging Percolation and Random Graphs: Phase transition in Dimension 1. Technical Report. arXiv:math.PR/0609594

C. T. Turova, T. Vallier: Merging Percolation on $\mathbb{Z}^d$ and Classical Random Graphs: Phase Transition. Under submission.

D. T. Vallier: Spread of Activation on Random Graphs. Technical Report.

Introduction

1 Motivation

The theory of random graphs has a wide range of applications in real life networks. It is a useful tool when studying the world wide web, neuron networks or social networks. To explain what a graph is, we can take a basic example of a social network like affinities in a class. We model each student as a vertex and if the students are friends then there is an edge between the two vertices assigned to the students to show the connection. When studying such a graph, questions arise like:
Are there any isolated student?
Are they all connected to anyone else through a path of friends or are there disconnected groups?
Who's the most popular student?
Is there a big group of friends that outnumber other groups?

These questions are part of the study of the structure of the graph. They all seem very simple but become much more complicated when one deals with large networks such as internet where there are billions of web pages linked in a large network. These questions can even become crucial in epidemiology studies, for example if one wants to know how to contain the spread of a virus with a limited stock of vaccine. Who should we vaccinate to stop the contamination?

2 Definition and terminology

Let us introduce some notations to study the structure of the graphs rigorously. We emphasise each definition with an example in real life networks to avoid a heavy accumulation of concepts without explanation of their purpose and illustrate definitions in figure 1.

Definition 1. *A graph G is an ordered pair of disjoint sets $G = (V, L)$ such that L, the set of edges is a subset of the set $V^2 = V \times V$ of unordered pairs of the set of vertices V.*

In the modelling of a network, the vertices often represent the entity and the edges are the connections or acquaintances if one thinks about a social network.

If u and v are two vertices of the graph then we denote the edge between these two vertices by (u, v). If $(u, v) \in L$ (i.e. we have an edge between the vertices u and v) then we say that the vertices u and v are *c*onnected through the edge (u, v). If a vertex w is connected to no other vertex, that is

$$w \in V, \forall v \in V, v \neq w, (w, v) \notin L$$

then it is *isolated*.

In modelling a network of acquaintance by a random graph as above, we assume that the fact to know someone is mutual. This assumption is not always true. One example is

the graph of the streets of a city. Some streets are two ways which corresponds to an edge and some are only one way. Therefore we must introduce the notion of orientation in a graph.

If the graph is *directed* (i.e. the edges are directed) then the pair $(\overrightarrow{u,v})$ denotes that there is an edge from the vertex u to the vertex v.

Consider that two vertices are connected through an edge if the persons modelled by those vertices are close enough to exchange germs. Someone ill might give the sickness to anyone with which it is connected. This person might in turn give the infection to its contacts and so on. If a virus appears in a very dense place then it's more likely to spread than in a scarcely inhabited one. This example allows us to introduce two important notions: the *connected component* and the *degree* of a vertex.We define the *connected components* as the sets of vertices linked by a chain of edges. An isolated vertex is the smallest possible component. The *largest connected component* is denoted C_1. If you are in a component disconnected from the original place of the infection then you can't be infected.

In a graph, the vertices can be linked to several other vertices. The vertices which are linked to u are called the neighbours of u and their set is denoted $\Gamma(u)$. The number of edges to which u is an endpoint is the degree of the vertex u which we denote

$$d(u) = \Big|\{(u,v), v \in V\}\Big|.$$

In a densely populated city you encounter more people and exchange more germs than in a unpopulated countryside. You are connected to more people which translates in terms of graphs as a vertex with a high degree.

In the case of directed graphs, we make a distinction between edges $(\overrightarrow{v,u})$ and $(\overrightarrow{u,v})$. Thus we make the same distinction by considering the in-degree $d_{in}(u)$ and the out-degree $d_{out}(u)$ of the vertex where the in-degree is the number of incoming edges

$$d_{in}(u) = \Big|\{(\overrightarrow{v,u}), v \in V\}\Big|$$

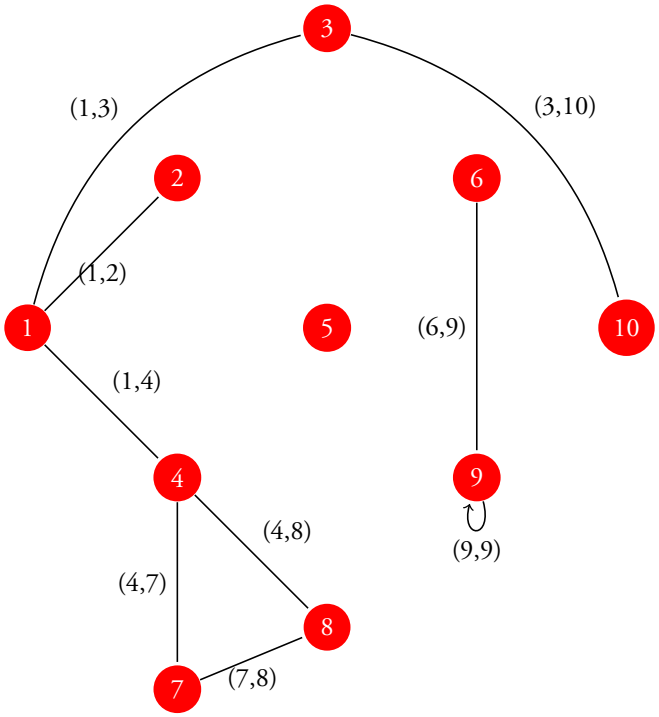
and the out-degree is the number of out-coming edges. We define the degree of a vertex u as

$$d(u) = d_{in}(u) + d_{out}(u)$$

In the figure 1 we consider an example of a non-directed graph to visualise the notions introduced in this section. Circles represent vertices $\big(V = \{1,2,...,10\}$ and the couple (u,v) where $u, v \in V$ indicates the edge connecting vertices u and v.

3 The degree sequence

The study of the degree sequence gives indications on the homogeneity of the graph. If the degrees are concentrated around a certain value then we can consider that the vertices



The neighbourhood of the vertex 1 is $\Gamma(1) = \{2, 3, 4\}$.
The degree of the vertex 1 is 3, $d(1) = 3$.
The vertex 5 is isolated.
There's a loop on vertex 9.
The vertices $\{1, 2, 3, 4, 7, 8, 10\}$ form a connected component.

Figure 1: An example of a graph.

are hardly distinguishable. If instead we have a very wide range of degrees then we will have highly connected vertices and not highly connected ones.

3.1 The degree sequence in classical random graph

The results given in this introduction about classical random graphs are taken from the book Random Graphs by Janson, Łuczak and Ruciński (2000) (**?**) and from the book with the same name by Bollobás (1985) (**?**).

The classical model of random graph denoted $G_{n,p}$ was introduced by Gilbert (1959) in (**?**) but it is commonly called Erdös-Rényi random graph since they first set the basis of the probabilistic treatment of $G_{n,p}$ in a series of papers in the 60's (see (**?**)).

Definition 2. *The model $G_{n,p}$ consists of all graphs with vertex set $V = \{1, ..., n\}$ in which all possible edges are chosen independently and with a probability p where $p \in [0, 1]$.*

Consider a vertex v_i which is connected to any vertex independently from the other connections with probability p. Let k_i denote the degree of the vertex v_i, the probability that the vertex v_i has degree k in the graph $G_{n,p}$ is binomial $(n-1, p)$

$$\mathbb{P}\{k_i = k\} = \binom{n-1}{k} p^k (1-p)^{n-1-k} \tag{3.1}$$

where p^k is the probability that the vertex has k edges,
$(1-p)^{n-1-k}$ is the probability of absence of further edges
and $\binom{n-1}{k}$ is the number of ways of selecting the neighbouring of v_i.

The events $\{k_i = k\}$ and $\{k_j = l\}$ are not independent. It is for example impossible that a vertex has a strictly positive degree while the others have degree 0. Thus the distribution of the number of vertices with degree k doesn't follow straightforward from (3.1).

Let X_k be the number of vertices of degree k in the graph then for a large range of probability p we have the following.

Theorem 3.1 (see (**?**))**.** *Let $\varepsilon > 0$ be fixed and let $\varepsilon n^{-\frac{3}{2}} \leq p = p(n) \leq 1 - \varepsilon n^{-\frac{3}{2}}$. Let $\lambda_k = n\binom{n-1}{k} p^k (1-p)^{n-1-k}$ then X_k has asymptotically a Poisson distribution with mean λ_k*

$$\mathbb{P}\{X_k = r\} \sim e^{-\lambda_k} \frac{\lambda_k^r}{r!} \tag{3.2}$$

for every fixed r, where $\sim$ means asymptotically distributed.

However, in real life networks we see a totally different type of behaviour. It has been found over the last decade that many different networks such as the citation patterns where each manuscript is a vertex and edges represent citations (i.e. there exists and edge directed from u to v if the manuscript u cites the manuscript v) have a power law

distribution of the degrees. In the real life networks, the proportion of vertices with degree k is best described as a power law:

$$\mathbb{P}(k) \sim k^{-\gamma}. \tag{3.3}$$

Redner (1998) (**?**) showed that $\gamma = 3$ for the network of citations. In the internet where a vertex represents a web page and the edges are links pointing from one page to another, the proportion of vertices with degree k follows a power law with $\gamma = 2,1 \pm 0,1$ (**?**).

3.2 Preferential attachment model

Recently, many new models of random graphs have been introduced motivated by the power law sequence explained above. Barabási and Albert (1999) (**?**) introduced the *preferential attachment model* based on two observations:

1. Most real world networks are open and continuously incorporate new vertices to the system. Think for example about internet and the new pages appearing on the web everyday.

2. Attachment is not uniform but is preferentially to vertices that already have a large number of connections. In the citation network mentioned above, a new manuscript is more likely to cite a well known and thus much cited paper. As an example, have a look at the bibliography of the introduction.

In this model, at each time step t a new vertex v_t is introduced with m edges linking v_t to the previous vertices with probabilities proportional to their degrees or attractiveness. Note that the same principle has been previously introduced in the little cited paper by Szymański (1987) (**?**).

Start with G^1, the graph with one vertex and one loop. Given G^{t-1} we form G^t by adding a vertex t together with a single edge $\left(\overrightarrow{t, g(t)}\right)$, where $g(t)$ is a random vertex chosen as follows. Let $D_i(t)$ denote the degree of vertex i at time t.

$$\mathbb{P}\Big(g(t) = i | D_i(t-1) = d_i\Big) = \begin{cases} \frac{d_i}{2t-1}, & \text{if } 1 \leq i \leq t-1\,, \\ \frac{1}{2t-1}, & \text{if } i = t\,. \end{cases}$$

Barabási, Albert and Jeong (1999) (**?**) showed by heuristic arguments supported by simulations that the degree distribution is proportional to d^{-3}. Rigorous proof of the power law distribution of degrees was given in (**?**). Here we give the Theorem where the number of edges introduced at each step is 1.

Theorem 3.2 (Bollobás, Riordan, Spencer, Tusnády (2001))**.** *Let $\sharp^n(d)$ denote the number of vertices with in-degree equal to d (i.e. with total degree $d+1$). Let*

$$\alpha(d) = \frac{4}{(d+1)(d+2)(d+3)}$$

and let $\varepsilon > 0$ *be fixed. Then with probability tending to* 1 *as* $n \to \infty$ *we have*

$$(1-\varepsilon)\alpha(d) \leq \frac{\sharp^n(d)}{n} \leq (1+\varepsilon)\alpha(d)$$

for every d *in the range* $0 \leq d \leq n^{\frac{1}{15}}$

By construction vertices with a high degree attract more edges and therefore become even more attractive. This phenomenom leads to a graph which accumulates most of the edges on the first vertices. The expected degree at time t of the vertex i introduced at time i is given by

$$\mathbb{E}\big(D_i(t)\big) = \prod_{j=i}^{t} \frac{2j}{2j-1} \sim \sqrt{t/i}.$$

This leads to a graph where the first vertices maintain the structure of the graph. Thus the model is robust against random deletion of vertices and edges but is vulnerable to an attack on the first vertices. This is explained by Bollobás and Riordan (2003) in (**?**). The introduction of the deletion has another purpose but just checking the strength of the network. In the preferential attachment model, vertices and edges are added at each time step but never deleted. However replacement is a natural rule in a network. A social network increases by the introduction of new arrivals but also evolves in a manner that connections can disappear. Using this premise, we consider the graph where an edge is deleted a time Δ after its introduction. We prove in paper **A** that for any fixed time Δ the expected degree is uniformly bounded by a constant $e^{\frac{1}{2}}$. This shows a phase transition at $\Delta = \infty$ and proves that the preferential attachment is not robust against the ageing of edges. Moreover we prove that for any fixed Δ, the degree of any vertex goes to 0 with time.

Similar deletion was studied by Turova in (**?**), (**?**), (**?**) and (**?**). Moreover, there the life time Δ of any edge is exponentially distributed. We consider in paper **A** a life time of any edge to be a constant just for mathematical tractability. However, it is clear that our model admits generalisations. Allowing some freedom on the deletion of edges would be a closer fit to reality. If for example, the deletion happens randomly on an interval of a fixed length δ centred on Δ, then our results are still valid. Consider now a few examples of real-world networks which topological or dynamical properties are similar to our model.

Biological networks

Our model fits the description of evolution where each vertex is a species and the in-degree quantifies the influence of the species in nature. Each species rise up to a climax and finally fade away until extinction. The growth phase refers to the period after each great extinction event when there's little or no competition between species until new species evolve better adaptations increasing come. Our model fulfils conditions of an evolution model stated by Newman (1996) in (**?**). "Constant change is a natural feature of evolution, on a sufficiently large scale in time, there's nothing remotely stable about evolution."

It is also noted that "one billion species have inhabited the planet since Cambrian ($\sim$ -500 millions years), only a few million are still living and most species become extinct about 10 millions years after their first appearance." Our model is in good agreement with that observation. It is also assumed (**?**) that "the number of species the ecosystem can sustain is roughly constant over time" which is verified in our model where the number of activated vertices for $m = 1$ is between $\frac{3}{2}\Delta$ and 2Δ (making the natural assumption that old vertices are connected by only one edge).

Neural networks

We may say that the vertex "dies" as soon as its degree becomes zero, and the entire graph is being renewed and evolves over time. Hence our model is also in good agreement with the models of neural networks (see, Iglesias et al. (2005) (**?**) for a relevant description and citations). Roughly, it is often assumed that the synaptic connection is lost if it was not activated by an impulse for some time. Also, in a neural network a new edge (impulse) from some vertex is created if this vertex receives enough energy from other vertices. This resembles a preferential attachment: a probability of sending out a new edge is increasing with in-degree.

Social Networks

Jin, Girvan and Newman (2001) (**?**) consider a model of social networks where acquaintances between pairs of individuals who rarely meet decay over time. They find an upper limit on the number of friendships (links) an individual can maintain. Our model agrees with this result: the links are deleted after a while and any vertex can maintain only a finite number of connections.

Our model can be applied to other networks evolving in time in which competition occurs. The same analysis can be also done for the model of Antal and Krapivsky (2005) (**?**) where edges carry a weight and the vertices' weight is the sum of the weight of the adjacent edges. Assigning the same weight to every edge, we recover the Preferential Attachment model. If we introduce deletion in this model, one can show that with time the weight and therefore the degree converges to 0 when time goes to infinity. If the weight assigned to newly introduced edges converges with time to a constant different from 0 then the degree is also uniformly bounded in Δ.

In the Growing Network by Copying model of Krapivsky and Redner (2005) (**?**), every new vertex attaches uniformly to a previously attached one and to those, the "target vertex" already points to. If one introduces deletion then the vertices incorporated before $t - \Delta$ have out-degree 0. The probability that the new vertex is connected to a vertex with out-degree 0 goes to one when t goes to infinity and the expected number of links goes to Δ. This incites us to add the condition that the target vertex has a strictly positive out-degree (which is not a condition with the inner model since all the vertices have out-degree strictly positive). With this assumption, the total number of links is increasing and is concentrated on the Δ last vertices. Hence, the maximal expected degree increases too in this model.

4 The largest connected component

The size of the largest connected component is a major study in the theory of random graphs. A simple application is to know whether or not a virus might spread to a large part of the population. Since the celebrated paper by Erdös and Rényi (1960) (**?**) the size of the largest connected component is well known in the classical random graph according to the probability p. We are mainly interested in the range when the size of the largest connected component abruptly *jumps* from a size of order $\log n$ to a positive part of the graph (order n).

4.1 The largest connected component in classical random graph

Consider a graph $G_{n,p}$. When $p = 0$ then with probability 1 there is no edge. The graph is totally disconnected and consists of isolated vertices. Conversely, if $p = 1$ then with a probability 1 any vertex is connected to all other vertices and the graph is fully connected. In the intermediary cases, we have very different structures depending on the value of p. In the study of random graphs, we consider p as a function of the number of vertices $p = p(n)$. When we increase p, the properties of the random graph change as the graph becomes denser in the sense that we have more edges. It is striking to see that the changing are sudden. The probability that a property holds can rapidly change from 0 to 1 as we increase $p(n)$. The range of the probability where this occurs is called the threshold function.

Definition 3. *(see (**?**)) We define the threshold function $f(n)$ of a property $\mathcal{P}$ by*

$$\mathbb{P}(\mathcal{P}) = \begin{cases} 0 \text{ if } \frac{p(n)}{f(n)} \to 0 \text{ as } n \to \infty \\ 1 \text{ if } \frac{p(n)}{f(n)} \to \infty \text{ as } n \to \infty \end{cases}$$

A monotone increasing property is a property that still holds by addition of new edges.

For every monotone property, the threshold function exists. As an example, the property that the graph is connected is an increasing property since addition of new edges cannot disconnect the graph. The size of the largest connected component can only increase by addition of new edges. Moreover, the addition of new edges can merge together two components into one big component. The increase of p implies the introduction of new edges which are more likely to link large components together than small ones. That way, the largest connected components increase their size. For $p(n)$ sufficiently large, the largest connected components merge into a giant component. This phase is called phase transition. The threshold function for the property that the largest connected component contains a positive part of the graph which corresponds to the phase transition is $f(n) = \frac{1}{n}$. This result has been proved in a Theorem by Erdös and Rényi (1960)

Theorem 4.1 (Erdös, Rényi (1960))**.** *Let $p(n) = \frac{c}{n}$, where $c > 0$ is a constant.*

- *If* $c < 1$ *then*

$$\lim_{n\to\infty} \mathbb{P}\Big\{|C_1| \leq \frac{1}{1-c-\log c}\log n\Big\} = 1.$$

- *If* $c > 1$, *let* $\beta(c) \in (0,1)$ *be uniquely defined by the equation*

$$\beta(c) + e^{-c\beta(c)} = 1. \tag{4.4}$$

Then for any $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\Big\{\Big|\frac{|C_1|}{n} - \beta(c)\Big| > \varepsilon\Big\} = 0.$$

See (**?**) for a proof of the supercritical case using a branching process argument.

4.2 Infinite component in percolation

The results from this section can be found in the book Percolation (1999) (**?**) by G. Grimmett. Conversely to the classical random graph, in percolation theory, the graph has a geometric structure. In *bond percolation theory*, the vertices are located on a square lattice and there is an open edge with the nearest neighbours with a probability p. In dimension 1 the vertices are ordered on a line and can have a link with the previous and the next vertices. In dimension d the neighbourhood is composed of the $2d$ vertices surrounding the vertex. To give an example in nature of an environment that can be nicely modelled by percolation, consider a volcanic stone that you immerge into water. Water goes through the holes modelled by open edges. Will the water reach the centre of the stone? We can consider the problem the other way around. We start from the centre and want to know with which probability there is a path of open edges to the borders. Considering an infinite lattice the question comes down to ask if with a positive probability there is an infinite connected cluster of open edges. In dimension 1, the critical probability above which there exists an infinite connected cluster is obviously $p_c = 1$ because if $p < 1$ then the probability that the cluster has a size larger than k decreases exponentially fast to 0. This problem is for higher dimension as simple to state as it is complex to study. Actually, only the critical probability in dimension 2 is exactly known $p_c(2) = \frac{1}{2}$ while for higher dimension, we can approximate the value of $p_c(d)$ through algorithm supported by computers. Denote the lattice in dimension d by $\mathbb{Z}^d$, we can restate the previous remark.

Theorem 4.2. *(see (**?**))*
The critical probability of bond percolation on $\mathbb{Z}^2$ *equals* $\frac{1}{2}$*:*

$$p_c(2) = \frac{1}{2}$$

If $p < p_c(2) = \frac{1}{2}$ then the lattice is composed by finite open clusters separated by an infinite closed cluster.
For $p = \frac{1}{2}$ then the probability of having an open and a closed edge is equal. The graph is composed of finite open and closed clusters.
If $p > \frac{1}{2}$ then conversely to the first case there is a infinite open cluster. Moreover, this cluster is with probability 1 unique.

Theorem 4.3. *(see (?))*
If $p > p_c$, then

$$\mathbb{P}_p(\textit{there exists exactly one infinite open cluster}) = 1.$$

The percolation transition is markedly similar to the phase transition in random graph. They both show a phase transition where the size of the largest connected component suddenly increases from something negligible with respect to the whole graph to a unique giant connected component which contains a positive part of the graph.

However, the models are very different in nature since there is no distance between the vertices in the classical random graph model while the geometry is fundamental in percolation. By not incorporating distance between vertices, the classical model misses properties of real world networks. On the other hand, the restriction to the nearest neighbours in percolation is not in agreement with observations. In a social networks for instance we are more likely to have acquaintance with neighbours or colleagues but we do not restrict our relations to such a narrow selection. A model that would capture the feature of both classical random graph model and percolation model would better model the connections in networks.

4.3 A model merging classical random graph and percolation

The model introduced in the paper **B** incorporates the classical model of random graphs and percolation. It is highly motivated by the model designed to study biological neural networks and introduced by Turova and Villa (2007) in (?) where vertices correspond to neurons which are connected by axons represented by edges. In papers **B** and **C**, we do not concentrate on the spread of the activation in the neuron networks but on the structure of the networks and more precisely on the phase transition. The exact formulation of the model is as follows.

We consider a graph on the set of vertices $V_N^d := \{1, \dots, N\}^d$ in $\mathbb{Z}^d$, where the edges between any two different vertices i and j are presented independently with probabilities

$$p_{ij} = \begin{cases} p, & \text{if } |i-j| = 1, \\ c/N^d, & \text{if } |i-j| > 1, \end{cases}$$

where $0 \leq p \leq 1$ and $0 < c < N$ are constants. This graph, call it $G_N^d(p,c)$ is a mixture of percolation model, where each pair of neighbours in $\mathbb{Z}^d$ is connected with probability

p, and a random graph model, where each vertex is connected to any other vertex with probability $\frac{c}{|N^d|}$.

This model can be seen as an attempt to generalise classical models into a single model. A generalisation of the classical random graph into an inhomogeneous random graph for which the classical model is a special case has been introduced and studied by Söderberg (2002) in (**?**), (**?**) and (**?**). It was later extended in the manuscript by Bollobás, Janson and Riordan (2007) (**?**).

We prove in paper **B** that in the case when the dimension of the lattice is 1 there is a phase transition along both parameters c related to the classical random graph model and p for the percolation model. Suppose that $0 \leq p \leq 1$ is fixed then there exists a critical c denoted by

$$c^{cr}(p) = \frac{1-p}{1+p} .$$

such that if $c < c^{cr}(p)$, we are in the subcritical case when the largest connected component denoted by $C_1\big(G\big)$ has a size of order $\log N$ with a probability tending to 1 as $N \to \infty$. If $c > c^{cr}(p)$ then we are in the supercritical case when $C_1\big(G\big)$ spans over a positive part of the graph and

$$\frac{|C_1\big(G_N(p,c)\big)|}{N} \xrightarrow{P} \beta$$

as $N \to \infty$, with $\beta = \beta(q,c)$ defined as the maximal solution to

$$\beta = 1 - \frac{1}{\mathbb{E}X}\,\mathbb{E}\left\{Xe^{-cX\beta}\right\}. \tag{4.5}$$

In paper **C**, we extend the results above to any dimension d. Let C denote an open cluster containing the origin of $\mathbb{Z}^d$ in the bond percolation model and $B(N)$ be the box of length N then

$$c^{cr}(p) = \frac{1}{\mathbb{E}|C|}.$$

If $c < c^{cr}(p)$ then with a probability tending to 1 as $n \to \infty$ we have $|C_1\big(G\big)| \leq \alpha \log|B(N)|$ with α given in paper **C**.
If $c \geq c^{cr}(p)$ then

$$\frac{|C_1\big(G_N(p,c)\big)|}{|B(N)|} \xrightarrow{P} \beta$$

as $N \to \infty$, with $\beta = \beta(p,c)$ defined as the maximal solution to

$$\beta = 1 - \mathbb{E}\left\{e^{-c\beta\,|C|}\right\}. \tag{4.6}$$

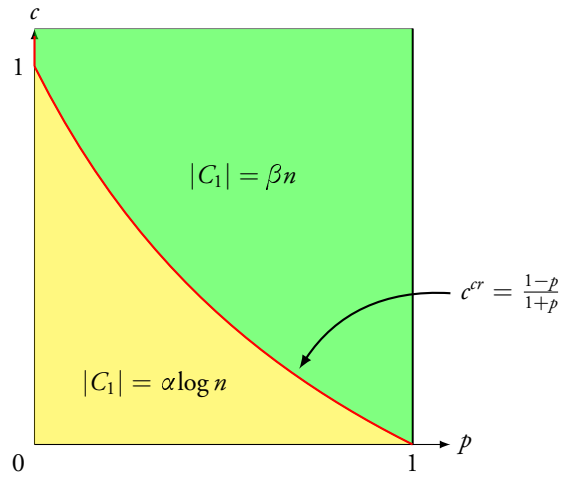


Figure 2: Phase diagram in dimension 1.

Even if given in terms of c^{cr} for p fixed, it is possible to state the same result for p^{cr} with a fixed c. The phase transition happens along both parameters. The duality of the parameter p and c is manifest in dimension 1(see figure 2), where

$$c^{cr}(p) = \frac{1-p}{1+p} \text{ and } p^{cr}(c) = \frac{1-c}{1+c}.$$

Notice that if we choose $p = 0$ then we do not have anymore influence from the percolation model and the model is equivalent to a classical random graph then equations (4.6) and (4.5) become (4.4).

This model is the combination of two homogeneous random graph models in the sense that the probability law of connection is the same for any vertex. It is itself an homogeneous random graph model. However, we use the theory of inhomogeneous random graph developed in (**?**) to tackle this problem. We consider the clusters formed by percolation (we know that for $p < p_c$ their size is finite) and consider each cluster as a macro-vertex. We build a graph on macro-vertices where each macro-vertex is of type k if the cluster contains k vertices and clusters are connected if there exists at least one edge between two vertices belonging to each macro-vertex. We derive the size of the giant connected component in equations (4.5) and (4.6) directly from (**?**) while the subcritical case requires an entire treatment.

5 Process of activation in random graphs

The last manuscript (paper **D**) of this thesis is dedicated to the spread of activation on a classical random graph. In the paper **B** and **C**, we studied the structure of the graph proposed in (**?**) to model neuron connections. Here we focus on the activation aspect and study the conditions under which the activation starting from a random set of activated vertices $A_n(0)$ at time 0 on a graph of n vertices spreads through the graph. In terms of biological network, it corresponds to an information that spreads in the neural network. It is known that a neuron needs several excitory impulses to become in turn excited. In paper **D**, we consider a simple process of activation where any vertex becomes activated if it has a link with at least two activated vertices. This model resembles a contact process and fits definition of a probabilistic cellular automata given in (**?**).

Definition 4. *Let A be a finite set of state. A probabilistic cellular automaton on $\mathbb{Z}^d$ is a stochastic process giving rise to a sequence of configurations*

$$\Phi_t : \mathbb{Z}^d \to A$$

where the state $\Phi_t(x)$ of $x \in \mathbb{Z}^d$ at time t is determined randomly with probabilities dependent on the states of the points of the neighbourhood of x at time $t-1$.

In our model, the neighbourhood as stated in the definition 4 is the entire set of vertices which can take two different states, activated or non-activated. A vertex changes from non-activated to activated with a probability depending on the probability of connection p of the vertices and the number of activated vertices.

We show that if the probability of connections is $p(n) = \frac{c}{n}$ where c is some constant and n is the number of vertices then the activation can't spread from a negligible part of the graph to a positive part of the graph. Taking $p(n)$ to be larger, we exhibit two different phases depending on the size of the original set of activated vertices $|A_n(0)|$. With probability tending to 1 as $n \to \infty$ we have the following

- If $|A_n(0)| = o(\frac{1}{np^2(n)})$ then the activation doesn't spread and the limiting set of activation has a size negligible with respect to n
- If for any $\nu > 0$, we have $|A_n(0)| > \frac{1+\nu}{np^2(n)}$ then the activation spreads through the entire graph.

Both the connections and the size of the inner set of activated vertices play a role in the transition.

This phenomenon is very natural. For instance, either you manage to contain virus or the infection expands faster and faster as the number of infected people increases. If the disease spreads to a positive part of the population, it is more likely that almost all the population will be infected leaving a few healthy people. This picture is fully described in the model of paper **D**. We summarise the results in the figure 3. This gives the size of

		$\lvert A_n(0)\rvert$			
		$o\left(\frac{1}{np^2(n)}\right)$	$\frac{1-\nu}{np^2(n)}$	$\frac{1+\nu}{np^2(n)}$	$\frac{\alpha(n)}{np^2(n)}$
$p(n)$	$\frac{c}{n}$	$o(n)$	†	†	$\surd$
	$\frac{c}{n} \ll \frac{\omega(n)}{n} \ll \frac{1}{\sqrt{n}}$	$o(n)$	?	$n\big(1-o(1)\big)$	$n\big(1-o(1)\big)$
	$\frac{c}{\sqrt{n}}$	$\surd$	$*n\big(1-o(1)\big)$		$n\big(1-o(1)\big)$
	$\frac{\omega(n)}{\sqrt{n}}$	$\surd$	$\surd$	$\surd$	$n\big(1-o(1)\big)$

Figure 3: The results are given with a probability tending to 1 as $n \to \infty$ except for the case marked with a $*$ where the event happens with a positive probability. In the cases marked with † we already have $|A_0(n)| > \varepsilon n$ for some $\varepsilon > 0$. The case marked with $\surd$ have no sense since either $|A_0(n)| = o(1)$ or $|A_0(n)| > n$.

the limiting set of activated vertices $A_n(n)$ with respect to the probability of connections $p(n)$ and the size of the inner set of activated vertices $A_n(0)$.

Notice that in the two cases marked with a †, we may have $|A_n(0)| > n$ depending on the constant c.

5.1 Further research

The properties of the model for $|A_n(0)| = \frac{1-\nu}{np^2(n)}(1 + o(1))$ with $0 \leq \nu < 1$ are not yet known. This implies a lack of knowledge on the type of phase transition. The analysis of the model can be extended to the case when the vertices need $k \geq 2$ connections with already activated vertices to become activated. From that, it would be possible to consider a model closer to the one exposed in (**?**).

Each vertex is given a random potential $X_v(0) \in [0, 1]$. A vertex is activated if its potential is 1. Start with a set of activated $A_n(0)$, each activated vertex sends a potential ω through the edges to its neighbours. This increases the potential of the targeted vertex

$$X_v(t) = \min\{1, X_v(0) + k\omega\}$$

where k is the number of activated vertices at time $t - 1$ which share an edge with the vertex v. The vertex v in turn becomes activated if its potential reaches 1. The study of such a process would be of great interest in modelisation of neuron networks.

Bibliography

[] Antal T.,Krapivsky P. L. (2005), Weight Driven Growing Networks *Phys. Rev E* **71** 026103.

[] Balister P., Bollobás B. and Kozma R. (2006), Large Deviations for Mean Field Models of Probabilistic Cellular Automata. *Random Structures and Algorithms* **29** no. 3, 399-415.

[] Barabási A. L., Albert R. (1999), Emergence of Scaling in Random Networks. *Science* **286**, 509-512.

[] Barabási A. L., Albert R., Jeong H. (1999), Mean-Field Theory for Scale-Free Random Networks. *Physica A* **272** 173-187.

[] Bollobás B.(1985), *Random Graphs.* Academic Press, London .

[] Bollobás B., Janson S. and Riordan O. (2007), The phase transition in inhomogeneous random graphs. *Random Structures and Algorithms* **31** 3-122.

[] Bollobás B., Riordan O., Spencer J., Tusnády G. (2001), The Degree Sequence of a Scale Free Random Graph Process. *Random Structures and Algorithms* **19**, 279-290.

[] Bollobás B., Riordan O. (2003), Robustness and Vulnerability of Scale-Free Random Graphs. *Internet Mathematics* **I 1**, 1-35

[] Erdös P. and Rényi A. (1960), On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5** 17-61.

[] Grimmett G. (1999), *Percolation* Springer-Verlag, Berlin,.

[] Gilbert E. N. (1959), Random Graphs. *Ann. Math. Statist.* **30** 1141-1144.

[] Iglesias J., Eriksson J., Grize F., Tomassini M., Villa A. (2005), Dynamics of Pruning in Simulated Large-Scale Spiking Neural Networks. *BioSystems* **79** 11-20.

[] Janson S., Łuczak T., and Ruciński A. (2000), *Random Graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York,.

[] Jin E. M., Girvan M., Newman M. E. J. (2001), The Structure of Growing Social Networks. *Phys. Rev E* **64**, 046132.

[] Krapivsky P. L., Redner S. (2005), Network Growth by Copying *Phys. Rev E* **71** 036118.

[] Newman M. E. J. (1996), Self Organized Criticality, Evolution and the Fossil Extinction Record, *Proceedings of the Royal Society of London* **B.263** 1605-1610.

[] Redner S. (1998), How Popular is Your Paper? An Empirical Study of the Citation Distribution. *European Physical Journal* **B 4**, 131-134.

[] Söderberg B. (2002), A General Formalism for Inhomogeneous Random Graphs, *Phys. Rev E* (3) **66**, 6, 066121, 6.

[] Söderberg B. (2002), Random Graphs with Hidden Color *Phys. Rev E* **68**, 2, 026107, 12.

[] Söderberg B. (2002), Properties of Random Graphs with Hidden Color. *Phys. Rev E* **68**, 2, 026107, 12.

[] Szymański J. (1987), On a Nonuniform Random Recursive Tree. *Annals of Discr. Math.* **33** 297-306.

[] Turova T. S. (2002), Dynamical Random Graphs with Memory. *Phys. Rev E* **65** 066102. (see also erratum *Phys. Rev E* **70** 059902 (2004).)

[] Turova T. S. (2003), Long Paths and Cycles in Dynamical Random Graphs. *J. Statist. Phys.* **110**, 1-2, 385-417.

[] Turova T. S. (2003), Phase Transitions in Dynamical Random Graphs. *J. Statist. Phys.* **123**, 5, 1007-1032.

[] Turova T. S. (2007), Continuity of the Percolation Threshold in Randomly Grown Graphs. *Electronic Journal of Probability* **12** 1036-1047.

[] Turova T. S. and Villa A. (2007), On a phase diagram for random neural networks with embedded spike timing dependent plasticity. *BioSystems* Volume 89, Issues 1-3, 280-286.