Università degli Studi di Padova

DIPARTIMENTO DI MATEMATICA 'TULLIO LEVI-CIVITA'
Corso di Laurea Triennale in Matematica

Tesi di Laurea

Boolean random networks: phase transition and connectivity

Candidato:
Giovanni Spisso
Matricola 2016100

Relatore:
Prof. Marco Formentin

Anno Accademico 2022-2023

Contents

Introduction

In a general and informal way, a spatial random network can be seen as a set of points randomly deployed on the plane and randomly connected to each other. Depending on the rules used to construct the network, its structure can resemble what is observed in real natural, as well as in artificial, complex systems.

Before starting, it is necessary to informally introduce the basic concepts that will be used throughout the thesis and the reason why random networks are relevant in many fields of science.

Historically, random networks have been studied in mathematics and statistical physics, although many models were inspired by practical questions of engineering interest. One of the first mathematical models was created in 1959 by Paul Erdös and Alfréd Rényi. They studied the behaviour of a 'typical' graph of n vertices where every edge connecting two vertices was drawn with probability p. They showed that many properties of these graphs are almost always predictable and they arise with very high probability when the model parameters are appropriately chosen. The graphs they considered, however, were abstract mathematical objects and there was no notion of geometric position of vertices and edges, which is fundamental for the models studied in this thesis.

Mathematical models inspired by more practical questions and relying on some notion of geometric locality of the random network connections appeared in 1957 in a paper published by Simon Broadbent and John Hammersley. They introduced a simple discrete mathematical model of a random grid in which vertices are arranged on a square lattice, and edges between neighbouring vertices are added at random. This simple model revealed an incredible mathematical depth and became one of the most studied mathematical objects in statistical physics. Broadbent and Hammersley were inspired by the work they had done during World War II and their paper's motivation was the optimal design of filters in gas masks. Indeed, the gas masks of the time used granules of activated charcoal, and they argued that the optimal functioning of the mask could be studied by simplifying the problem through a random network model. They studied what should be the correct density of the charcoal: at one extreme air flowed easily through the cannister, but the wearer of the mask breathed insufficiently filtered air, at the other extreme, the charcoal pack was nearly impermeable, and while no poisonous gases got through, neither did sufficient air. The optimum was to have high charcoal surface area and tortuous paths for air flow, ensuring sufficient time and contact to absorb the toxin. They realised that

this condition would be met in a critical operating regime, which would occur with very high probability. They named the mathematical framework that they developed percolation theory, because the meandering paths reminded them of water trickling through a coffee percolator.

A few years later, in 1961, American communication engineer Edgar Gilbert introduced a model of random planar networks in continuum space. He considered nodes randomly located in the plane and formed a random network by connecting pairs of nodes that are sufficiently close to each other. He wanted to use this model to provide long-range radio connection using a large number of short-range radio transmitters, and this was the starting point of continuum percolation theory. Using this model, he formally proved the existence of a critical transmission range for the nodes.

These networks can be used for human communication, as well as for sensing the environment and collecting and exchanging data for a variety of applications, such as environmental and habitat monitoring, industrial process control, security and surveillance, and structural health monitoring.

The object of this thesis is to study some fundamental properties and results of a particular type of random network, which is called the Boolean model.

In Chapter 1 we formally introduce the model and we explain how it can be constructed by probabilistic methods.

In Chapter 2 we analyse the presence of phase transition in the case of an infinite network. The purpose of this part of the thesis is to prove that there exists a critical density of the nodes for which the network suddenly changes its behaviour.

Finally, in Chapter 3, we study the connectivity of finite networks growing in size. We focus specifically on studying if the behaviour of the networks asymptotically corresponds to the one observed in the second chapter and on how the density of the nodes should grow in comparison to the size of the network for having all the nodes connected to each other almost surely.

At the end, we verify the results by doing some simulations.

All the main results discussed in this thesis are taken from [3].

Chapter 1

The boolean model

1.1 Poisson processes

As we said in the Introduction, there are several types of random networks, each one with its own properties. In particular, we want to study the Boolean model, a continuum model where the position of the nodes of the network themselves are random and are formed by the realisation of a *point process* on the plane. To properly define the Boolean model it is necessary to introduce some preliminary notions.

We start by listing the following desirable features of a somehow regular, random network deployment.

(i) *Stationarity*. The distribution of the nodes in a given region of the plane should be be invariant under any translation of the region to another location of the plane.

(ii) *Independence*. The number of nodes deployed in disjoint regions of the plane should be independent.

(iii) *Absence of accumulation*. The number of nodes in every bounded region of the plane should be finite and on average proportional to the area of that region.

We now describe a way to construct a process that has all the features listed above and later give its formal definition. Consider a square of side length one. Partition this square into n^2 identical subsquares of side length $1/n$ and assume that the probability p that a subsquare contains exactly one point is proportional to the area of the subsquare, so that for some $\lambda > 0$,

$$p = \frac{\lambda}{n^2}. \tag{1.1}$$

We assume that having two or more points in a subsquare is impossible. We also assume that points are placed independently of each other. Lets look at

the probability that the (random) number of points N in the whole unit square is k. This number of points is given by the sum of n^2 independent random variables, each of which has a small probability λ/n^2 of being equal to one, and which are equal to zero otherwise. It is well known that, as $n \to \infty$ this sum converges to the Poisson distribution of parameter λ, which is usually called *law of rare events*. Indeed,

$$\begin{aligned} \lim_{n\to\infty} P(N=k) &= \lim_{n\to\infty} \binom{n^2}{k} \left(\frac{\lambda}{n^2}\right)^k \left(1-\frac{\lambda}{n^2}\right)^{n^2-k} \\ &= \lim_{n\to\infty} \frac{\lambda^k}{k!} \left(1-\frac{\lambda}{n^2}\right)^{n^2} \frac{n^2!}{n^{2k}(n^2-k)!} \left(1-\frac{\lambda}{n^2}\right)^{-k} \\ &= \frac{\lambda^k}{k!} e^{-\lambda}. \end{aligned} \tag{1.2}$$

The construction in the unit square clearly satisfies the three desired properties, and we now want to extend it to the whole plane. Consider two disjoint unit squares and look for the distribution of the number of points inside them. This is the sum of two independent Poisson random variables, that is a Poisson random variable of parameter 2λ. This leads to the idea that in the point process on the plane, the number of points in any given region A should have a Poisson distribution of parameter $\lambda|A|$, where $|\cdot|$ denotes area. This intuition leads to the following definition.

Definition 1.1. (Poisson process) A random set of points $X \subset \mathbb{R}^2$ is said to be a *Poisson process* of density $\lambda > 0$ on the plane if it satisfies the conditions:

(i) For mutually disjoint domains of $\mathbb{R}^2$ $D_1, \ldots, D_k$, the random variables $X(D_1), \ldots, X(D_k)$ are mutually independent, where $X(D)$ denotes the random number of points of X inside domain D.

(ii) For any bounded domain $D \subset \mathbb{R}^2$ we have that for every $k \geq 0$:

$$P(X(D)=k) = e^{-\lambda|D|} \frac{(\lambda|D|)^k}{k!} \tag{1.3}$$

The definition does not explicitly say how the points are distributed on the plane so we now introduce a constructive procedure to create a Poisson process, which will also be useful to make the simulations. Lets start with the following observation.

Observation 1.1. Let $B \subset A$ be bounded sets. By conditioning on the number

of points inside $|A|$, and applying Definition 1.1, we have

$$\begin{aligned} P(X(B)=m|X(A)=m+k) &= \frac{P(X(B)=m, X(A)=m+k)}{P(X(A)=m+k)} \\ &= \frac{P(X(A\backslash B)=k, X(B)=m)}{P(X(A)=m+k)} \\ &\overset{(i)}{=} \frac{P(X(A\backslash B)=k)P(X(B)=m)}{P(X(A)=m+k)} \\ &\overset{(ii)}{=} \binom{m+k}{m}\left(\frac{|A|-|B|}{|A|}\right)^k\left(\frac{|B|}{|A|}\right)^m. \end{aligned} \tag{1.4}$$

This expression is a binomial distribution with parameters $m+k$ and $|B|/|A|$. If we condition on the number of points in a region A to be $m+k$, then the number of points in $B \subset A$ represent the number of successes in $m+k$ experiments with success probability $|B|/|A|$. This means that each of the $m+k$ points is randomly and uniformly distributed on A and the positions of the different points are independent. This proves that the property is necessary, we will now show that it is also sufficient, namely a process that satisfies this property is actually a Poisson process. Hence, to construct a Poisson point process in any bounded region A of the plane we should do the following: first draw a random number N of points from a Poisson distribution of parameter $\lambda|A|$, and then distribute these uniformly and independently over A.

Proposition 1.2. *Let $N, M_1, \ldots, M_r$ be random variables with the following properties:*

(i) *N has a Poisson distribution with parameter μ;*

(ii) *The conditional distribution of the vector $(M_1, \ldots, M_r)$ given $N = s$ is multinomial with parameters s and $p_1, \ldots, p_r$.*

Under these conditions $M_1, \ldots, M_r$ are mutually independent Poisson distributed random variables with parameters $\mu p_1, \ldots, \mu p_r$ respectively.

Proof. Define $m_1 + \cdots + m_r = s$:

$$\begin{aligned} P(M_1=m_1,\ldots,M_r=m_r) &= P(M_1=m_1,\ldots,M_r=m_r|N=s)P(N=s) \\ &= \frac{s!}{m_1!\cdots m_r!}p_1^{m_1}\cdots p_r^{m_r}e^{-\mu}\frac{\mu^s}{s!} \\ &= \prod_{i=1}^{r}\frac{p_i^{m_i}}{m_i!}e^{-\mu p_i}. \end{aligned}$$

□

N represents the number of points in A in the construction above, and $M_1, \ldots, M_r$ represent the number of points ending up in regions $B_1, \ldots, B_r$ into which A has been subdivided. Since the properties of the construction are now translated into properties (i) and (ii) of the proposition above, the conclusion is that the number of points in disjoint regions are mutually independent with

the correct Poisson distribution. Hence, this really is a constructive procedure to create a Poisson process on A.

Finally, note that the independence property of the process implies that the process is not affected by conditioning on the event that there is a point at $x_0 \in \mathbb{R}^2$, apart from that point. Indeed, we can for example divide the plane into infinitesimal regions and, for the independence property, the presence of the point in one region does not influence the others. This fact can be stated with arbitrarily high level of formality, we refer to [9] for the technical details.

The definition of a Poisson point process can be generalized to the case when the density is not constant over the plane, but it is a function of the position over $\mathbb{R}^2$. We will use this type of process to prove some theorems later in the thesis.

Definition 1.3. (Inhomogeneous Poisson process) A countable set of points $X \subset \mathbb{R}^2$ is said to be an *inhomogeneous Poisson process* on the plane with density function $\Lambda : \mathbb{R}^2 \to [0, \infty)$, if it satisfies the conditions

(i) For mutually disjoint domains of $\mathbb{R}^2$ $D_1, \ldots, D_k$, the random variables $X(D_1), \ldots, X(D_k)$ are mutually independent, where $X(D)$ denotes the random number of points inside domain D.

(ii) For any bounded domain $D \subset \mathbb{R}^2$ we have that for every $k \geq 0$

$$P(X(D) = k) = e^{-\int_D \Lambda(x)dx} \frac{[\int_D \Lambda(x)dx]^k}{k!} \tag{1.5}$$

Proposition 1.4. *Let X be a Poisson point process with density λ on the plane, and let $g : \mathbb{R}^2 \to [0, 1]$. Consider a realization of X and delete each point x with probability $1 - g(x)$, and leave it where it is with probability $g(x)$, independently of all other points of X. This procedure generates an inhomogeneous Poisson point process of density function $\lambda g(x)$.*

Proof. We denote by $\tilde{X}$ the point process after having applied the procedure in the proposition. The independence property is immediate from the construction, and the distribution of $\tilde{X}$ can be computed as follows:

$$P(\tilde{X}(A) = k) = \sum_{i=k}^{\infty} P(X(A) = i) P(\tilde{X}(A) = k | X(A) = i). \tag{1.6}$$

Given the event $X(A) = i$, the i points of X in A are uniformly distributed over A because of (1.4). Thus the conditional distribution of $\tilde{X}$ given $X(A) = k$ is just

$$P(\tilde{X}(A) = 1 | X(A) = 1) = |A|^{-1} \int_A g(x)dx, \tag{1.7}$$

and more generally,

$$\begin{aligned} P(\tilde{X}(A) = k | X(A) = i) = &\binom{i}{k} \left(|A|^{-1} \int_A g(x)dx \right)^k \\ &\times \left(1 - |A|^{-1} \int_A g(x)dx \right)^{i-k}. \end{aligned} \tag{1.8}$$

Hence,

$$\begin{aligned} P(\tilde{X}(A)=k) &= e^{-\lambda|A|}\frac{(\lambda\int_A g(x)dx)^k}{k!} \\ &\quad\times\sum_{i=1}^{k}\frac{(\lambda|A|[1-|A|^{-1}\int_A g(x)dx])^{i-k}}{(i-k)!} \\ &= e^{-\lambda|A|}\frac{(\lambda\int_A g(x)dx)^k}{k!}e^{\lambda|A|(1-|A|^{-1}\int_A g(x)dx)} \\ &= \frac{(\lambda\int_A g(x)dx)^k}{k!}e^{-\lambda\int_A g(x)dx}. \end{aligned}$$

□

1.2 Boolean networks

We start by giving a general definition that includes the one of boolean networks and that will be important in the next paragraphs.

Definition 1.5. (Poisson random connection networks) A *Poisson random connection model*, denoted by (X,λ,g), is given by a Poisson point process X of density $\lambda>0$ on the plane, and a connection function $g(\cdot):\mathbb{R}^2\to[0,1]$ satisfying the condition $0<\int_{\mathbb{R}^2} g(x)dx<\infty$.

Each pair of points $x,y\in X$ is connected by an edge with probability $g(x-y)$, independently of all other pairs and of X. We also assume that $g(x)$ depends only on the Euclidean norm $|x|$ and is non-increasing in the norm.

That is,

$$g(x)\le g(y) \text{ whenever } |x|\ge|y|.$$

This gives a type of network where the density λ plays a key role, as densely packed nodes form very different structures than sparse nodes.

The random connection model is quite general and has applications in different branches of science. In physics the random connection function may represent the probability of formation of bonds in particle systems; in epidemiology the probability that an infected herd at location x infects another herd at location y; in telecommunications the probability that two transmitters are non-shaded and can exchange messages; in biology the probability that two cells can sense each other.

We now give the proper definition of boolean network.

Definition 1.6. (Boolean network) For a given $r>0$, a *Boolean model network* is a Poisson random connection model where the connection function is of the boolean zero-one type,

$$g(x)=\begin{cases}1 & \text{se } |x|\le 2r\\ 0 & \text{se } |x|>2r\end{cases} \tag{1.9}$$

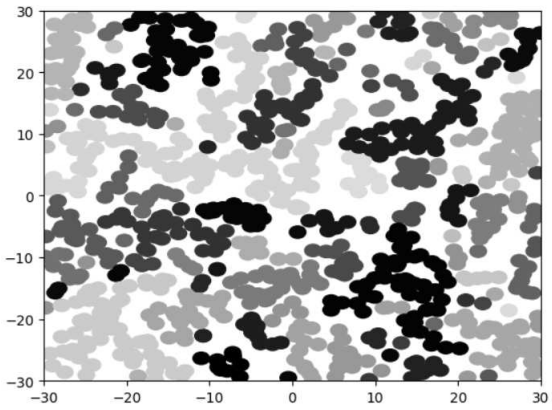

Figure 1.1: Boolean model. Connected components of overlapping discs are drawn with the same grey level.

This geometrically corresponds to placing discs of radius r at the Poisson points and considering connected components formed by clusters of overlapping discs; see Figure 1.1.

The *Boolean model* also has applications in many fields. In communication engineering it can be adopted as a first-order approximation of communication by isotropic radiating signals. This applies to radio communication, and more generally to any communication where signals isotropically diffuse in a noisy environment. In [2], for example, the authors analyse the applications of the *Boolean model* to wireless communications. Even some biological systems communicate by diffusing and sensing chemicals. We refer, as an example, to [10], an article studying the usefulness of the *Boolean model* in physical biology. Another possible field where this model can be applied is the one of neural networks, as studied in [5].

Chapter 2

Phase transition

2.1 Random connection model

Informally, a *phase transition* is defined as a phenomenon by which a small change in the *local* parameters of a system results in a huge change of its *global* behaviour, which can be observed over an infinite domain.

We now consider the random connection model introduced in section 1.2. In the following, we always condition on a Poisson point being at the origin. This is possible for the independence property as said in the previous chapter.

It is easy to see that the integrability condition is required to avoid a trivial model. Indeed, let Y denote the (random) number of points that are directly connected to the origin. This number is given by an inhomogeneous Poisson point process of density $\lambda g(x)$, so that

$$P(Y = k) = e^{-\lambda \int_{\mathbb{R}^2} g(x)dx} \frac{[\lambda \int_{\mathbb{R}^2} g(x)dx]^k}{k!}, \tag{2.1}$$

where this expression is to be interpreted as zero in the case $\int_{\mathbb{R}^2} g(x)dx = \infty$. It follows that if $\int_{\mathbb{R}^2} g(x)dx = 0$, then $P(Y = 0) = 1$ and each point is isolated a.s. On the other hand, if $\int_{\mathbb{R}^2} g(x)dx$ diverges, then $P(Y = k) = 0$ for all finite k and in that case $Y = \infty$ a.s.

The following theorem is crucial because it shows that in the random connection model there is always a critical value of λ for which there is a phase transition. We define the number of vertices in the component at the origin as $|C|$ and $\theta(\lambda) = P_\lambda(|C| = \infty)$.

Theorem 2.1. *There exists a* $0 < \lambda_c < \infty$ *such that* $\theta(\lambda) = 0$ *for* $\lambda < \lambda_c$ *and* $\theta(\lambda) > 0$ *for* $\lambda > \lambda_c$.

Since in the proof of 2.1 are used some theorems that regard other particular types of random networks, we need to state a couple of preliminary observations and definitions.

Observation 2.1. $P_\lambda(|C| = \infty) > 0$ is equivalent to the existence a.s. of an unbounded connected component on $\mathbb{R}^2$.

Observation 2.2.

$$E(|C|) = \infty P(|C| = \infty) + \sum_{n=1}^{\infty} nP(|C| = n), \tag{2.2}$$

where $0 \times \infty$ is defined as 0. From (2.2), it follows that

(i) $E(|C|) < \infty$ implies $P(|C| = \infty) = 0$;

(ii) $P(|C| = \infty) > 0$ implies $E(|C|) = \infty$;

(iii) $E(|C|) = \infty$ implies nothing.

Definition 2.2. (Branching process) A branching process is obtained by a *tree* T composed of an infinite number of vertices, where each vertex has exactly $k > 0$ children, and draw each edge of the tree with probability $p > 0$, or delete it otherwise, independently of all other edges.

Definition 2.3. (Site percolation model) Consider an infinite square lattice $\mathbb{Z}^2$ and divide it in boxes. In a *site percolation model* of probability p, each box is occupied with probability p and empty otherwise, independently of all other boxes.

Proof of theorem 2.1. In the first part of the proof we prove that $\theta(\lambda) = 0$ by showing that $E(|C|) < \infty$, while in the second part, because of Observation 2.2, we use that $\theta(\lambda) > 0$ implies $E(|C|) = \infty$.

First of all we need to show the monotonicity of the percolation function $\theta(\lambda)$. To do so, we consider two random connection models with $\lambda_1 < \lambda_2$ and we thin the process of density λ_2, namely we delete each point of this model independently with probability $(1 - \lambda_1/\lambda_2)$. The resulting graph can be effectively viewed as a realisation of a model of density λ_1. On the other hand, it is also clear that the latter realisation contains less points than the original one, so we can conclude that if there is an infinite cluster in the realisation of density λ_1, then there must also exist one in the realisation of density λ_2, and this concludes the proof of the monotonicity of $\theta(\lambda)$.

We now prove the first part of the theorem. We start with a point x_0 in the origin and we consider points directly connected with x_0. We obtain an inhomogeneous Poisson point process of density $\lambda g(x - x_0)$, that we name the first generation. We denote these random points by $x_1, x_2, \ldots, x_n$ ordered by modulus. We continue considering the random points directly connected with x_1 but not with x_0. They also form an inhomogeneous Poisson point process, independent from the first and with density $\lambda g(x-x_1)(1-g(x-x_0))$. We go on as obvious, so that at each point x_i we form an independent inhomogeneous Poisson point process with the points still not connected with the previous ones, adding at the density a factor $(1-g(x-x_j))$ with $j = i-1$. This construction produces a random graph G such that if two points are connected with the origin in the

original random model, they are also connected in G. The number of points in the n-th generation has as an upper bound the number of points of the n-th generation of a random tree of expecting offspring $\mu = \lambda \int_{\mathbb{R}^2} g(x - x_n)dx = \lambda \int_{\mathbb{R}^2} g(x)dx$. That is because some connections in the construction are missing due to the additional factors $(1 - g(x - x_j)) < 1$. It is possible to choose λ small enough such that $\mu \leq 1$ and, for a fundamental theorem in random networks theory, we know that when $\mu \leq 1$ the branching process does not grow forever with probability one. We refer to [3] for the precise statement and proof of that theorem. This proves that $E(|C|) < \infty$, concluding the first part of the proof.

For the second part of the proof we need to prove that for λ big enough, $\theta(\lambda) > 0$. It is useful to define $\bar{g} : \mathbb{R}^+ \to [0, 1]$ by

$$\bar{g}(|x|) = g(x), \tag{2.3}$$

for all $x \in \mathbb{R}^2$.

Partition the plane into subsquares of side length one. The probability that two Poisson points inside two adjacent boxes are connected by an edge is at least $\bar{g}(\sqrt{5})$, since the diagonal of a rectangle formed by two adjacent boxes is $\sqrt{5}$. Moreover, the probability that at least k points are inside a box can be made larger than $1 - \epsilon$, for arbitrarily small ϵ, taking λ big enough. Take two adjacent boxes and let x_0 be a point of the Poisson process in one of them. For λ big enough, the probability that x_0 is connected to at least a point in the adjacent box is given by

$$p \geq (1 - \epsilon) \left(1 - \left(1 - \bar{g}(\sqrt{5}) \right)^k \right). \tag{2.4}$$

We choose k and λ such that $p > p_c$, the critical probability for the site percolation on the square lattice. In fact, another fundamental theorem in random networks theory states that also in the *site percolation model* there is phase transition, so p_c exists. The precise statement and proof of this theorem are also in [3].

Lets now describe a dynamic procedure similar to the one above that ensures percolation in the random connection model. Begin with a point $x_0 \in X$. In the first iteration determine the connections from x_0 to Poisson points in each of the four boxes adjacent to the one where x_0 is placed. We call each of these boxes occupied if there is at least a connection from x_0 to some point inside the box, empty otherwise. Note that everyone of these boxes is occupied, independently from the others, with probability $p > p_c$. In the second iteration move to a point x_1 inside an occupied box directly connected with x_0 and examine the connections from x_1 to points in boxes still not examined. This procedure goes on in the natural way each time determining the state of new boxes and spanning a tree rooted in x_0 that is a subgraph of the component centred in x_0 in the random connection model. Since $p > p_c$, the probability that the box of x_0 is in an unbounded connected component of adjacent boxes is positive, and this implies that x_0 is in an unbounded connected component of the random connection model with positive probability. □

2.2 Boolean model

We now state a theorem that is a directed consequence of Theorem 2.1 in the special case of the Boolean random network model.

Theorem 2.4.

(i) *In a boolean random network of radius r, exists a critical density* $0 < \lambda_c < \infty$ *such that* $\theta(\lambda) = 0$ *for* $\lambda < \lambda_c$*, and* $\theta(\lambda) > 0$ *for* $\lambda > \lambda_c$*.*

(ii) *In a boolean random network of density* λ*, exists a critical radius* $0 < r_c < \infty$ *such that* $\theta(r) = 0$ *for* $r < r_c$*, and* $\theta(r) > 0$ *for* $r > r_c$*.*

(iii) *In a boolean random network exists a critical degree* $0 < \xi_c < \infty$ *such that* $\theta(\xi) = 0$ *for* $\xi < \xi_c$*, and* $\theta(\xi) > 0$ *for* $\xi > \xi_c$*.*

The exact values of the critical quantities in the Theorem 2.4, are not known, but there are some analytic bounds that can be obtained adapting the proof of Theorem 2.1. We can, for example, substitute in the proof the connection function $g(x)$ with the indicator function obtaining some explicit bounds. Computer simulations suggest that $\xi_c = 4\pi r_c^2 \lambda_c \approx 4.512$.

The proof of Theorem 2.4 follows immediately from Theorem 2.1 and the following proposition.

Proposition 2.5. *In a boolean random network it is the case that*

$$\lambda_c(r) = \frac{\lambda_c(1)}{r^2}. \tag{2.5}$$

Proof. Consider a realisation G of the boolean network with $r = 1$. Scale all distances in this realisation by a factor r, obtaining a scaled network G_s. G_s can be seen as the realisation of a boolean model where all discs have radius $1/r$ and the density of the Poisson process is $\lambda(1)/r^2$. The connections of G and G_s are the same and this means that if G percolates, G_s also does so. On the other hand if G does not percolate, G_s does not either. It follows that $\lambda_c(G_s) = \lambda_c(1)/r^2$, which concludes the proof. □

Proof of Theorem 2.4. The point (i) has already been proven in Theorem 2.1 since the boolean model is a particular case of the Poisson random connection model.

For the point (ii) use Proposition 2.5 obtaining that the critical radius r_c is $\sqrt{\lambda_c(1)/\lambda_c(r)}$.

For the point (iii) combine the previous two points obtaining the critical node degree $\xi_c = 4\pi r_c^2 \lambda_c$. □

We now analyse the compression phenomenon. In every Poisson random connection model at high density, finite clusters tend to be formed by isolated points and they are also the last finite components to disappear as $\lambda \to \infty$. In the case of the boolean model is possible to study with precision this phenomenon.

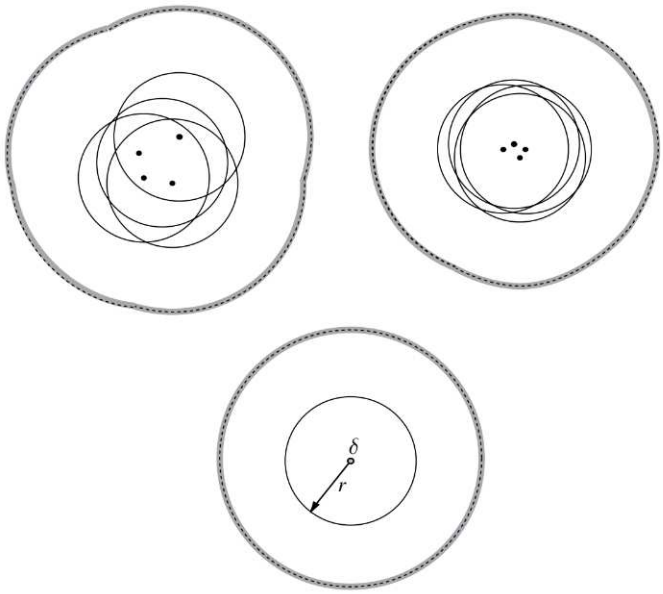

Figure 2.1: The compression phenomenon. *Image taken from* [3]*, p.49.*

For λ big enough, $P(|C| = k)$ is clearly very small for any fixed $k \geq 0$. More precisely a necessary and sufficient condition for this event to occur is the existence of a component of k connected discs surrounded by a fence of empty region not covered by discs. As we can see in 2.1, this is equal to have the region inside the highlighted dashed line not containing any Poisson point other than the given k, forming the isolated cluster of discs. Clearly, the area of this region is smaller when the k points are near. Hence, in a boolean model at high density the event $|C| = k$ is rare, but if it occurs, it is more likely in a configuration where the k points collapse into a very small region, and the approximately circular-shaped area around them of radius $2r$ is free of Poisson points.

Lets now do more precise considerations on the compression phenomenon.

Theorem 2.6. *In a boolean model with high density, the points tend to be isolated or part of an infinite connected component, more precisely,*

$$\lim_{\lambda \to \infty} \frac{1 - \theta(\lambda)}{e^{-\lambda \pi (2r)^2}} = 1. \tag{2.6}$$

Proof. We consider the following sufficient condition to have an isolated component of $k + 1 < \infty$ points: impose the presence of a point at the origin and name, for $\alpha < r$, $S = S_\alpha$ the event that a disc of radius α contains k additional points, and an annulus outside it with width $2r$ does not contain any Poisson process; see Figure 2.2. If S occurs, then there is an isolated cluster of size $k+1$ at the origin. The probability of the event S can be computed as follows:

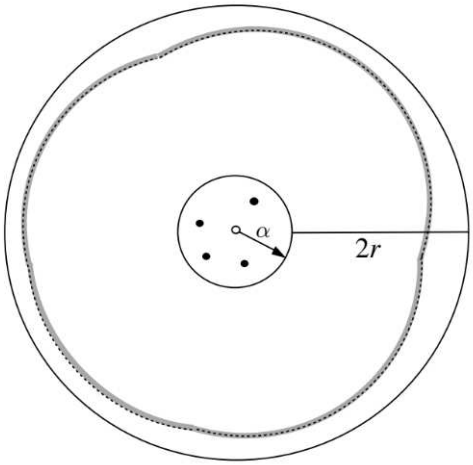

Figure 2.2: Sufficient condition for an isolated component. *Image taken from* [3]*, p.50.*

$$\begin{aligned} P(S) &= \frac{(\lambda\pi\alpha^2)^k}{k!} e^{-\lambda\pi\alpha^2} e^{-\lambda(\pi(\alpha+2r)^2-\pi\alpha^2)} \\ &= \frac{(\lambda\pi\alpha^2)^k}{k!} e^{-\lambda\pi(\alpha+2r)^2}. \end{aligned} \tag{2.7}$$

Since S implies the presence of a cluster of $k+1$ points at the origin, we also have

$$P(|C| = k+1) \geq P(S), \quad \text{for all } \alpha, \lambda, k. \tag{2.8}$$

We want to use $P(S)$ to approximate $P(|C| = k+1)$. To improve the approximation, we first maximise the lower bound over α,

$$P(|C| = k+1) \geq \max_{\alpha} P(S_\alpha), \quad \text{for all } \lambda, k. \tag{2.9}$$

Since we are interested in the behaviour at high density, we take the limit for $\lambda \to \infty$,

$$\lim_{\lambda\to\infty} \frac{P(|C| = k+1)}{\max_\alpha P(S_\alpha)} \geq 1, \quad \text{for all } k. \tag{2.10}$$

We now compute the denominator in (2.10). To find the maximum, derive in function of α (2.7), impose the first derivate equal to zero, control what points are in the domain (remembering that $0 < \alpha < r$) and finally verify what point maximizes $P(S)$, comparing the points also with the extremes 0 and r that could be maximum. After some counts, the maximum is

$$\alpha = \frac{k}{2\pi r\lambda} + O\left(\frac{1}{\lambda^2}\right). \tag{2.11}$$

For $k \neq 0$ we rewrite (2.10) as

$$\lim_{\lambda \to \infty} \frac{P(|C| = k+1)}{\exp(-\lambda\pi(2r)^2 - k \log \frac{\lambda}{k} - O(1))} \geq 1, \quad \text{for every } k. \tag{2.12}$$

Note that for $k = 0$ the approximation is sharp, in that case the denominator of (2.10) behaves as $e^{-\lambda\pi(2r)^2}$, which is exactly the probability that a node is isolated and appears also in the numerator. In fact the approximation is sharp for every value of k and to understand why, it is necessary to return to the sufficient condition S. In the equation (2.11), α tends to zero as $\lambda \to \infty$ and the annulus described in the definition of S becomes a disc of radius $2r$. But remembering that the disc of radius α must contain $k+1$ points, this implies that the points must be arbitrarily near each other for $\alpha \to 0$. Remember what happens to the *necessary* condition for k points to be isolated when they become near to each other: the empty region requested for the finite component to be isolated is exactly a disc of radius $2r$. It follows that when $\alpha \to 0$, the sufficient and the necessary condition to have $|C| = k+1$ are the same condition.

Finally, studying the denominator in (2.12), note that the probability to have limited components tends to zero faster for big values of k. This explains why isolated nodes are the last ones to disappear as $\lambda \to \infty$. Substituting $k = 0$ concludes the proof. □

Chapter 3

Connectivity in the boolean model

3.1 Preliminaries

Until now, we have discussed phase transition on infinite networks. We now want to study the asymptotic behaviour of sequences of finite random networks that grow larger in size. This topic interests scientists because real systems are of finite size and discovering the correct scaling laws that govern their behaviour, means discovering how the system is likely to behave as its size increases. In particular the main question that we want to answer is if the phase transition is a property present also in finite networks and in which sense there can be phase transition.

We start with a preliminary observation: there are two equivalent ways to create networks of a growing number of nodes. One can either fix the area of the network, and increase the density of the nodes to infinity or keep the density constant and increase the area of interest to infinity. Although the two cases above can describe different practical scenarios, by appropriate scaling of the distance lengths, they can be viewed as the same network realization.

Before we start stating the main theorems, we need to briefly introduce a powerful mathematical tool that we will frequently use in this chapter: the *Chen–Stein method*, which is used to estimate convergence to a Poisson distribution. We have already seen how a Poisson distribution naturally arises as the limiting distribution of the sum of n independent, low probability, indicator random variables. The idea behind the Chen–Stein method is that this situation generalizes to dependent, low probability random variables, as long as dependencies are negligible as n tends to infinity.

We first have to define a distance between two probability distributions.

Definition 3.1. The total variation distance between two probability distribu-

tions p and q on $\mathbb{N}$ is defined by

$$d_{TV}(p,q) = \sup\{|p(A) - q(A)| : A \subset \mathbb{N}\} \tag{3.1}$$

We need to introduce the following notation. Let I be an arbitrary index set, and for $\alpha \in I$, let I_α be an indicator random variable with expectation $E(I_\alpha) = p_\alpha$. We define

$$\lambda = \sum_{\alpha \in I} p_\alpha \tag{3.2}$$

and assume that $\lambda < \infty$. Let $W = \sum_{\alpha \in I} I_\alpha$, and note that $E(W) = \lambda$. Finally, we denote with $Po(\lambda)$ a Poisson random variable with parameter λ.

We now state a bound on the total variation distance between the Poisson distribution of parameter λ and the distribution of the sum of n dependent indicator random variables with expectations p_α. We do not prove the bound, which appears in [1]. It also makes use of the notion of *neighbourhood of dependence* as defined below.

Definition 3.2. For each $\alpha \in I$, $B_\alpha \subset I$ is a neighbourhood of dependence for α, if I_α is independent of all indices I_β, for $\beta \notin B_\alpha$.

Theorem 3.3. *Let B_α be a neighbourhood of dependence for $\alpha \in I$. Let*

$$\begin{aligned} b_1 &\equiv \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} E(I_\alpha) E(I_\beta), \\ b_2 &\equiv \sum_{\alpha \in I} \sum_{\beta \in B_\alpha, \beta \neq \alpha} E(I_\alpha I_\beta). \end{aligned} \tag{3.3}$$

It is the case that

$$d_{TV}(W; Po(\lambda)) \leq 2(b_1 + b_2). \tag{3.4}$$

We use this bound in the particular case of indicator random variables of events in random networks whose probability decays with n. In this case the bounds converge to zero and the sum of the indicators converges in distribution to the Poisson distribution with parameter $\lambda = E(W)$.

We now formally introduce the notions necessary to study finite boolean models. Let X be a Poisson process of unit density on the plane. We consider the boolean random network model $(X, \lambda = 1, r > 0)$. As in the previous chapter we condition on a Poisson point being at the origin and let $\theta(r)$ be the probability that the origin is in an infinite connected component. We focus on the restriction $G_n(r)$ of the network formed by the vertices that are inside a box $B_n \subset \mathbb{R}^2$ of side $\sqrt{n}$. We call $N_\infty(B_n)$ the number of Poisson points in B_n that are part of an infinite connected component in the boolean model $(X, 1, r)$ over the whole plane. As said before, all the results obtained also hold, by scaling, considering a box of unit length, density $\lambda = n$ and dividing all distance lengths by $\sqrt{n}$. We start by proving the following proposition.

Proposition 3.4. *We have $\theta(r) = E[N_\infty(B_1)]$.*

Proof. Divide B_1 into m^2 subsquares $s_i, i = 1, \ldots, m^2$ of side length $1/m$ and define a random variable X_i^m that has value one if there is exactly one Poisson point in s_i, that is also contained in an infinite connected component of the whole plane, and zero otherwise. Let $X_m = \sum_{i=1}^{m^2} X_i^m$. X_m is a non decreasing sequence that tends to $N_\infty(B_1)$ as $m \to \infty$. Hence, by the monotone convergence theorem, we also have that

$$\lim_{m\to\infty} E(X_m) = E(N_\infty(B_1)). \tag{3.5}$$

We now call s_i *full* if it contains exactly one Poisson point, and let A_i be the event that a Poisson point in s_i is part of an infinite component. Finally, call $\theta_m(r)$ the conditional probability $P(A_i|s_i \text{ full})$.

We note that $\theta_m(r) \to \theta(r)$ for $m \to \infty$. This is true because if $m \to \infty$ and s_i is full, s_i tends to become equivalent to a point of the Poisson process. The probability of the point being in an infinite component is the same as the origin because of the stationary property present in the definition of the Poisson process.

We have

$$\begin{aligned} E(X_i^m) &= P(X_i^m = 1) \\ &= P(A_i|s_i \text{ full})P(s_i \text{ full}) \\ &= \theta_m(r)\left[\frac{1}{m^2} + o\left(\frac{1}{m^2}\right)\right]. \end{aligned} \tag{3.6}$$

It follows that

$$E(X_m) = m^2 E(X_i^m) = [1 + o(1)]\theta_m(r). \tag{3.7}$$

By taking the limit for $m \to \infty$ in (3.7) and using (3.5), we obtain

$$E(N_\infty(B_1)) = \lim_{m\to\infty}[1 + o(1)]\theta_m(r) = \theta(r). \tag{3.8}$$

□

We know from the previous chapter that in a Boolean model defined on the whole plane, the percolation function $\theta(r)$ represents the probability that a single point is in an infinite connected component. One might expect that the fraction of the points in the main cluster inside the box B_n is roughly equal to this function. This would mean that also in the case of finite networks we have a phase transition and that there is a value of the radius of the discs that allows a certain fraction of the nodes in B_n to be connected. On the other hand, if one wants to observe all nodes to be connected inside the box, then the radius of the discs must grow with the box size. We make these considerations precise below, starting with almost connectivity.

3.2 Almost connectivity

Definition 3.5. For any $\alpha \in (0, 1)$, $G_n(r)$ is said to be α-*almost connected* if it contains a connected component of at least αn vertices.

The main theorem regarding almost connectivity in the boolean model is the following.

Theorem 3.6. *Let*

$$r_\alpha = \inf\{r; \theta(r) > \alpha\}. \tag{3.9}$$

We have that for any $\alpha \in (0,1)$, if $r > r_\alpha$, then $G_n(r)$ is α-almost connected asymptotically a.s., while for $r < r_\alpha$ it is not.

Before proving Theorem 3.6 we need to mention a preliminary result.

Theorem 3.7. *Consider a supercritical boolean model of radius r and density $\lambda > \lambda_c$. For any $0 < \delta < 1$, let $R_{\delta n}$ be a rectangle of sides $\sqrt{n} \cdot \delta\sqrt{n}$ on the plane. Let $R^{\leftrightarrow}_{\delta n}$ denote the event of a left to right crossing inside the rectangle, that is, the existence of a connected component of Poisson points of $R_{\delta n}$, such that each of the two smaller sides of $R_{\delta n}$ has at least a point of the component within distance r from it. We have*

$$\lim_{n\to\infty} P(R^{\leftrightarrow}_{\delta n}) = 1. \tag{3.10}$$

Proof of Theorem 3.6. The proof is based on geometric constructions. We begin proving that for $r > r_\alpha$, $G_n(r)$ is α-almost connected. A sufficient condition to have a connected component in $G_n(r)$ containing αn vertices is the presence of a box $B_{\delta n}$ containing at least αn points of an infinite connected component (we call this event A_1) surrounded by a circuit $G_n(r)$ (we call this event A_2); see Figure 3.1. We will prove that each of these events holds with an arbitrarily high probability as $n \to \infty$. The result immediately follows from the union bound.

$$\begin{aligned} P(A_1 \cup A_2) &= 1 - P(A_1^c \cap A_2^c) \\ &\geq 1 - P(A_1^c) - P(A_2^c) \to 1 \end{aligned} \tag{3.11}$$

for $n \to \infty$.

Lets begin by looking for a circuit of $G_n(r)$ surrounding $B_{\delta n}$. By Theorem 3.7 we have that if $r > r_c$, for any $0 < \delta < 1$, there exists a crossing path in a rectangle of sides $\sqrt{n} \times \sqrt{n}(1 - \sqrt{\delta}/2)$ with high probability. We apply this result to the four rectangles that surround $B_{\delta n}$, as we can see in the figure 3.2.

We call CR_i, $i \in \{1,2,3,4\}$ the four events that denote the existence of crossing paths inside everyone of the four rectangles, CR^i their complements.

By the union bound we have

$$\begin{aligned} P\left(\bigcap_{i=1}^{4} CR_i\right) &= 1 - P\left(\bigcup_{i=1}^{4} CR_i^c\right) \\ &\geq 1 - \sum_{i=1}^{4} P(CR_i^c) \to 1, \end{aligned} \tag{3.12}$$

as $n \to \infty$.

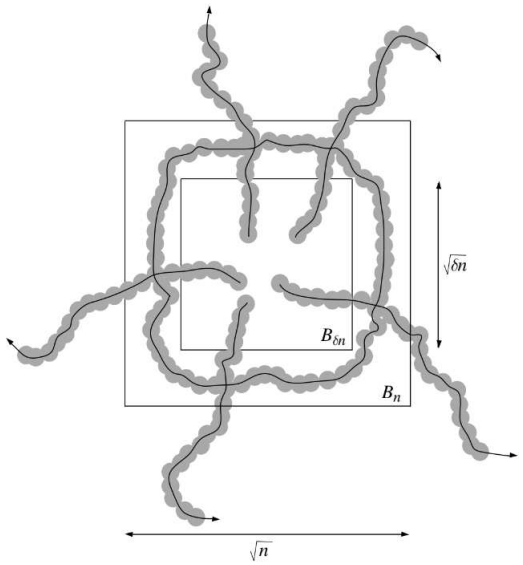

Figure 3.1: Sufficient condition for almost connectivity. *Image taken from* [3], *p.79.*

The next step is to show that for all $0 < \alpha < 1$ there are at least αn points inside $B_{\delta n}$ that are part of an infinite connected component of the boolean model on the plane. We choose $r > r_\alpha$ so that $\theta(r) > \alpha$. Then, using Proposition (3.4), we can choose $0 < \delta < 1$ and $\epsilon > 0$ such that

$$\delta E[N_\infty(B_1)] = \delta\theta(r) \geq \alpha + \epsilon. \tag{3.13}$$

From (3.13) follows that

$$\begin{aligned} P(N_\infty(B_{\delta n}) < \alpha n) &= P\left(\frac{N_\infty(B_{\delta n})}{n} < \alpha\right) \\ &\leq P\left(\left|\frac{N_\infty(B_{\delta n})}{n} - \delta E[N_\infty(B_1)]\right| > \epsilon\right). \end{aligned} \tag{3.14}$$

To conclude this part of the proof we need to use the *Ergodic Theorem*, a generalisation of the classical strong law of large numbers (SLLN). The classical SLLN states that the average of many i.i.d. random variables tends to their common expectation. However, in this case the random variables are not independent, but they are stationary for the property (i) characterizing a Poisson process.

The *Ergodic Theorem* is exactly what we need since it generalises the SLLN to stationary identically distributed random variables. For more details on the *Ergodic Theorem* we refer to [6].

It follows that we have, a.s.,

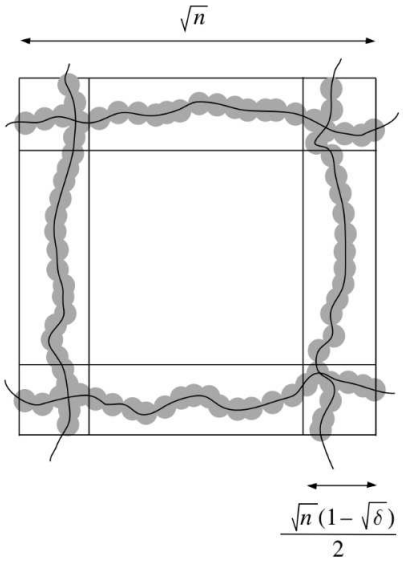

Figure 3.2: Existence of the circuit. *Image taken from* [3]*, p.79.*

$$\lim_{n\to\infty} \frac{N_\infty(B_{\delta n})}{\delta n} = E(N_\infty(B_1)). \tag{3.15}$$

Since a.s. convergence implies convergence in probability, it follows that the right-hand side of (3.14) tends to zero for $n \to \infty$, which completes the first part of the proof.

We now need to prove that if $r_c < r < r_\alpha$, then less than $\alpha = \theta(r_\alpha)$ nodes are connected. To do this, partition B_n into M^2 subsquares s_i of side length $\sqrt{n}/M$ for some fixed $M > \sqrt{4/\alpha}$. Let $\delta \in (1-\alpha/4, 1)$, and let w_i be the square of area $\delta|s_i| < \delta n\alpha/4$, placed at the centre of s_i and A_i the annulus $s_i \setminus w_i$; see Figure 3.3.

Note that with these definitions,

$$\frac{|w_i|}{|s_i|} = \delta > 1 - \frac{\alpha}{4} \tag{3.16}$$

and hence

$$\frac{|A_i|}{|s_i|} < \frac{\alpha}{4}. \tag{3.17}$$

Finally, we also have that $|s_i| < n\alpha/4$.

We consider the following events.

(i) Every s_i, $i = 1, \ldots, M^2$, contains at most $\alpha n/4$ vertices.

(ii) $\cup_{i=1}^{M^2} A_i$ contains at most $\alpha n/4$ vertices.

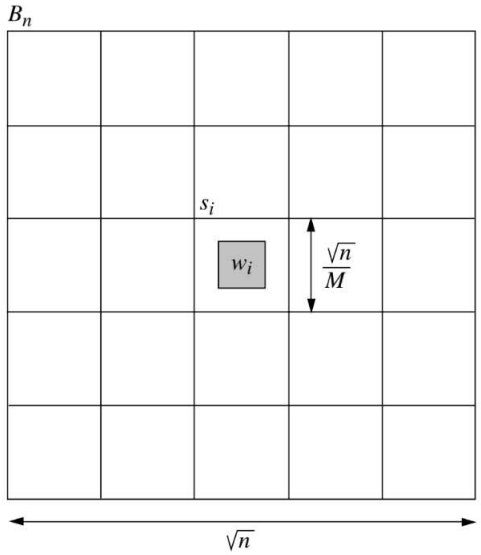

Figure 3.3: Partition of the box and annuli construction. *Image taken from* [3], *p.81.*

(iii) $N_\infty(B_n) < \alpha n$.

(iv) All annuli A_i, $i = 1, \ldots, M^2$, contain circuits that are part of the unbounded component.

We now analyse the probability of these events. The ergodic theorem tells us that the number of points in a large region deviates from its mean by at most a small multiplicative factor. Hence, the event in (i) occurs w.h.p. since $|s_i| < n\alpha/4$. The event (ii) also occurs w.h.p. since the union of the annuli A_i cover less than a fraction $\alpha/4$ of the square. At the same way we obtain the result for event (iii). To conclude, event (iv) also occurs w.h.p. by the argument used in the first part of the proof.

We claim that the occurrence of events (i)-(iv) implies that no component in B_n can have more than αn vertices. This is true because each component that has vertices in two boxes w_i and w_j, $i \neq j$, also connects to the circuits in A_i and A_j that are part of an infinite component. This implies by (iii) that it contains less than αn vertices. It remains to rule out the possibility of having components of size at least αn that are contained in $s_j \cup \bigcup_{i=1}^{M^2} A_i$, for some $j = 1, \ldots, M^2$. But by (i) and (ii) the number of vertices of this set is at most $2\alpha n/4 < \alpha n$ and this completes the proof. □

3.3 Full connectivity

We now want to study how the density λ should grow, with respect to the size of the network, for having *all* points inside the box B_n forming a connected cluster. We have just demonstrate that α-connectivity is achieved above the critical percolation radius r_α. The intuition in that case was that above criticality the infinite component invades the whole plane, including the area inside the box, and makes a fraction of the nodes in the box connected. This fraction is asymptotically equal to the value $\theta(r)$ of the percolation function. Now, if we want to observe a fully connected cluster inside a growing box, we clearly need to grow the radius of the discs with the box size. The problem is to identify at what rate this must be done. In the following, we see what is the exact threshold rate for asymptotic connectivity. We begin with a preliminary result that shows the required order of growth of the radius.

Theorem 3.8. *Let $\pi r_n^2 = \alpha \log n$. If $\alpha > 5\pi/4$, then $G_n(r)$ is connected w.h.p., while for $\alpha < 1/8$ it is not connected w.h.p.*

Proof. We first show that $G_n(r)$ is not connected for $\alpha < 1/8$. Consider two concentric discs of radii r_n and $3r_n$ and let A^n be the event that there is at least one Poisson point inside the inner disc and there are no Poisson points inside the annulus between radii r_n and $3r_n$. We have that

$$\begin{aligned} P(A^n) &= (1 - e^{-\pi r_n^2}) e^{-8\pi r_n^2} \\ &= \left(\frac{1}{n}\right)^{8\alpha} \left[1 - \left(\frac{1}{n}\right)\right]^{\alpha}, \end{aligned} \tag{3.18}$$

where we have used that $\pi r_n^2 = \alpha \log n$. Consider now 'packing' the box B_n with non-intersecting discs of radii $3r_n$. There are at least $\beta n/(\log n)$ of such discs that fit inside B_n, for some $\beta > 0$. A sufficient condition to avoid full connectivity of $G_n(r)$ is that A^n occurs inside at least one of these discs. Accordingly,

$$P(G_n(r) \text{ not connected}) \geq 1 - (1 - P(A^n))^{\frac{\beta n}{\log n}}. \tag{3.19}$$

By (3.18) and exploiting the inequality $1 - p \leq e^{-p}$ that holds for any $p \in [0, 1]$, we have

$$(1 - P(A^n))^{\frac{\beta n}{\log n}} \leq \exp\left[-\frac{\beta n}{n^{8\alpha} \log n}\left(1 - \left(\frac{1}{n}\right)^{\alpha}\right)\right], \tag{3.20}$$

which converges to zero for $\alpha < 1/8$. This completes the first part of the proof.

We now need to show that $G_n(r)$ is connected w.h.p. for $\alpha > 5\pi/4$. Partition B_n into subsquares S_i of area $\log n - \epsilon_n$, where $\epsilon_n > 0$ is chosen so that the partition is composed of an integer number $k = n/(\log n - \epsilon_n)$ of subsquares

and $n/\log n = k + x$, with $0 < x < 1$. We have

$$\begin{aligned} \epsilon_n &= \log n - \frac{n}{k} \\ &= \frac{x \log n}{k} \\ &< x \frac{(\log n)^2}{n - \log n} \end{aligned} \tag{3.21}$$

We call a subsquare *full* if it contains at least one Poisson point, and we call it *empty* otherwise. The probability for a subsquare to be empty is $e^{-logn+\epsilon_n}$, while the probability that every subsquare of B_n is full is

$$P\left(\bigcap_{i=1}^{\frac{n}{\log n - \epsilon_n}} S_i \text{ is full}\right) = (1 - e^{-\log n + \epsilon_n})^{\frac{n}{\log n - \epsilon_n}}. \tag{3.22}$$

Note that this latter probability tends to one as $n \to \infty$ since $\epsilon_n = o(1)$ from 3.21. Note also that any two points in adjacent subsquares are separated by at most a distance of $(5 \log n - 5\epsilon_n)^{1/2}$, which is the length of the diagonal of the rectangle formed by two adjacent subsquares. It follows that if

$$r_n > \frac{\sqrt{5 \log n - 5\epsilon_n}}{2}, \tag{3.23}$$

then every point in a subsquare connects to all points in that subsquare and also to all points in all adjacent subsquares. This is the same condition as

$$\pi r_n^2 > \frac{\pi 5}{4}(\log n - \epsilon_n). \tag{3.24}$$

By dividing both sides of the inequality in (3.24) by $\log n$ and taking the limit for $n \to \infty$, it follows that for $\alpha > 5\pi/4$, points in adjacent subsquares are connected. Since by (3.22), w.h.p. every subsquare contains at least a Poisson point, the result follows. □

The following Theorem gives a stronger result regarding the precise rate of growth of the radius to obtain full connectivity and it is the main result of this section.

Theorem 3.9. *Let $\pi(2r_n)^2 = \log n + \alpha_n$. Then $G_n(r_n)$ is connected w.h.p. if and only if $\alpha_n \to \infty$.*

Idea of the proof of Theorem 3.9. A complete proof of this theorem is quite technical and long. We do not attempt to give all details here, however, we want to highlight the main steps required to obtain a rigorous proof.

The first step is to show that isolated nodes do not arise w.h.p. inside the box if and only if $\alpha_n \to \infty$. This first step is shown by Proposition 3.10 and Proposition 3.11 below.

The second step is to show that ruling out the possibility of having isolated nodes inside the box is equivalent to achieving full connectivity of all nodes inside the box. To do this, first we state in Theorem 3.12 that the longest edge of the nearest neighbour graph among the nodes in B_n has the same asymptotic behaviour as the longest edge of the tree connecting all nodes in B_n with minimum total length. Then, by Proposition 3.13 we show that this is also the same asymptotic behaviour of the critical radius for full connectivity of the boolean model inside the box.

The key to the first step is to approximate the sum of many low probability events, namely the events that a given node is isolated, by a Poisson distribution. One complication that arises in this case is given by boundary conditions. It is in principle possible that isolated nodes are low probability events close to the centre of the box, but that we can observe 'fake singletons' near the boundary of it. These are Poisson points that are connected on the infinite plane, but appear as singletons inside the box.

The key to the second step is that at high density (or at large radii), if the cluster at the origin is finite, it is likely to be a singleton; then simply ruling out the possibility of observing isolated points inside a finite box should be sufficient to achieve full connectivity. However, even if we show that singletons cannot be observed anywhere in the box, and we know by the compression phenomenon that when radii are large no other isolated clusters can form, it is in principle possible to observe extremely large clusters that are not connected inside the box, but again only through paths outside the box. Theorem 2.6 simply does not forbid this possibility. Hence, the step from ruling out the presence of singletons inside the box to achieving full connectivity is not immediate. Finally, note that the compression theorem focuses only on the cluster at the origin, while we are interested in all points inside the box. To adapt this theorem to a finite box and ensure that all we observe is likely to be a singleton, the probability of being a singleton conditioned on being in a component of constant size, must converge to one sufficiently fast when we consider the union of all points inside the box. All of these difficulties are carefully overcome in the work of Penrose [7], and in the following we only give an outline of this work.

Proposition 3.10. *If $\pi(2r_n)^2 = \log n + \alpha$, then the number of isolated nodes inside B_n converges in distribution to a Poisson random variable of parameter $\lambda = e^{-\alpha}$.*

We now state a slight variation of Proposition 3.10 that can be proven following the arguments in [3]. Note that this also shows that w.h.p. there are no isolated nodes inside B_n if and only if $\alpha_n \to \infty$.

Proposition 3.11. *Let $\pi(2r_n)^2 = \log n + \alpha_n$, and let A_n be the probability that there are no isolated nodes in B_n. We have that*

$$\lim_{n\to\infty} P(A_n) = e^{-e^{-\alpha}} \tag{3.25}$$

if and only if $\alpha_n \to \alpha$, where α can be infinity.

We now give a proof of Proposition 3.10 in the simpler case when B_n is a torus. This implies that we do not have special cases occurring near the boundary of the box, and the events inside B_n do not depend on the particular location inside the box.

Proof of Proposition 3.10 (Torus case). The proof is based on a discretisation of the space, followed by the evaluation of the limiting behaviour of the event that a node is isolated.

Partition B_n into m^2 subsquares centred in $s_i \in \mathbb{R}^2$, $i = 1, \ldots, m^2$ of side length $\sqrt{n}/m$, and denote these subsquares by V_i, $i = 1, \ldots, m^2$. Let A_i^{mn} be the event that V_i contains exactly one Poisson point. For any fixed n, and any sequence $i_1, i_2, \ldots$, we have

$$\lim_{m\to\infty} \frac{P(A_{i_m}^{mn})}{n/m^2} = 1. \tag{3.26}$$

Note that for fixed m and n, the events A_i^{mn} are independent of each other, and that the limit above does not depend on the particular sequence (i_m).

It is time to turn to node isolation events. Let D_n be a disc of radius $2r_n$ such that $\pi(2r_n)^2 = \log n + \alpha$, centred at s_i. We call B_i^{mn} the event that the region of all subsquares intersecting $D_n \backslash V_i$ does not contain any Poisson point. For any fixed n, and any sequence $i_1, i_2, \ldots$, we have

$$\lim_{m\to\infty} \frac{P(B_{i_m}^{mn})}{e^{-\pi(2r_n)^2}} = 1. \tag{3.27}$$

Note that in (3.27) the limit does not depend on the particular sequence (i_m), because of the torus assumption. Note also that events B_i^{mn} are certainly independent of each other for boxes V_i centred at points s_i further than $5r_n$ apart, because in this case the corresponding discs D_n only intersect disjoint subsquares.

We define the following random variables for $i = 1, \ldots, m^2$:

$$I_i^{mn} = \begin{cases} 1 \text{ if } A_i^{mn} \text{ and } B_i^{mn} \text{ occur,} \\ 0 \text{ otherwise,} \end{cases} \tag{3.28}$$

$$W_n^m = \sum_{i=1}^{m^2} I_i^{mn}, \quad W_n = \lim_{m\to\infty} W_n^m. \tag{3.29}$$

Note that W_n indicates the number of isolated nodes in B_n. We now want to use the Chen-Stein bound in Theorem 3.3. Accordingly, we define a neighbourhood of dependence N_i for each $i \le m^2$ as

$$N_i = \{j : |s_i - s_j| \le 5r_n\}. \tag{3.30}$$

Note that I_i^{mn} is independent of I_j^{mn} for all indices j outside the neighbourhood

of independence of i. Writing I_i for I_i^{mn} and I_j for I_j^{mn}, we also define

$$\begin{aligned} b_1 &\equiv \sum_{i=1}^{m^2} \sum_{j \in N_i} E(I_i)E(I_j), \\ b_2 &\equiv \sum_{i=1}^{m^2} \sum_{j \in N_i, j \neq i} E(I_i I_j). \end{aligned} \tag{3.31}$$

By Theorem 3.3 we have that

$$d_{TV}(W_n^m; Po(\lambda)) \leq 2(b_1 + b_2), \tag{3.32}$$

where $\lambda = E(W_n^m)$. Writing $a_m \sim_m b_m$ if $a_m/b_m \to 1$ as $m \to \infty$, using (3.26) and (3.27) we have

$$\begin{aligned} \lambda = E(W_n^m) &\sim_m n e^{-\pi(2r_n)^2} \\ &= e^{\log n - \pi(2r_n)^2} \\ &= e^{-\alpha}. \end{aligned} \tag{3.33}$$

Since the above result does not depend on n, we also have that

$$\lim_{n\to\infty} \lim_{m\to\infty} E(W_n^m) = \lim_{n\to\infty} e^{-\alpha} = e^{-\alpha}. \tag{3.34}$$

We now compute the right-hand side of (3.32). From (3.26) and (3.27) we have that

$$E(I_i) \sim_m \frac{n}{m^2} e^{-\pi(2r_n)^2}. \tag{3.35}$$

From this it follows that

$$\begin{aligned} \lim_{m\to\infty} b_1 &= \lim_{m\to\infty} \sum_{i=1}^{m^2} \left(\frac{n}{m^2} e^{-\pi(2r_n)^2} \right)^2 \frac{\pi(5r_n)^2}{n} m^2 \\ &= e^{-2\alpha} \frac{\pi(5r_n)^2}{n}, \end{aligned} \tag{3.36}$$

which tends to 0 as $n \to \infty$.

We want to show similar behaviour for b_2. The first thing to notice is that $E(I_i I_j)$ is zero if two discs of radius $2r_n$, centred at s_i and s_j, cover each other's centres, because in this case the event A_i^{mn} cannot occur simultaneously with B_j^{mn}. Hence, we have

$$E(I_i I_j) = \begin{cases} 0 & \text{if } 2r_n > |s_i - s_j| \\ P(I_i = 1, I_j = 1) & \text{if } 2r_n < |s_i - s_j|. \end{cases} \tag{3.37}$$

We now look at the second possibility in (3.37). Let $D(r_n, x)$ be the area of the union of two discs of radius $2r_n$ with centres a distance x apart. Since B_i^{mn} and

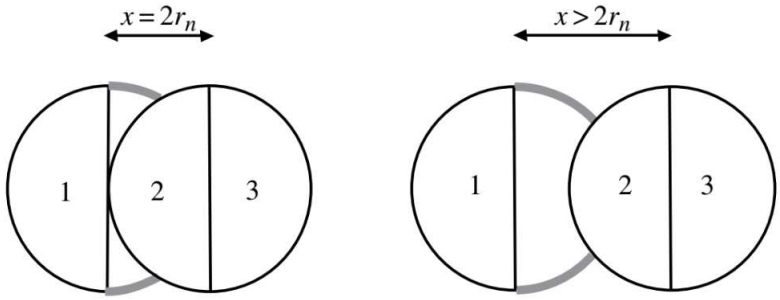

Figure 3.4: The union discs of radius $2r_n$ separated by a distance of at least $2r_n$ has area at least $3\pi(2r_n)^2/2$. *Image taken from* [3] *p.86.*

B_j^{mn} describe a region without Poisson points that tends to $D(r_n, |s_i - s_j|)$ as $m \to \infty$, for $2r_n < |s_i - s_j|$ we can write

$$E(I_i I_j) \sim_m \left(\frac{n}{m^2}\right)^2 \exp[-D(r_n, |s_i - s_j|)]. \tag{3.38}$$

We define annular neighbourhood $\mathcal{A}_i$ for each $i \leq m^2$ as

$$\mathcal{A}_i = \{j : 2r_n \leq |s_i - s_j| \leq 5r_n\}. \tag{3.39}$$

Combining (3.31), (3.37) and (3.38) we have

$$\begin{aligned}
\lim_{m\to\infty} b_2 &= \lim_{m\to\infty} \sum_{i=1}^{m^2} \sum_{j\in\mathcal{A}_i, j\neq i} \left(\frac{n}{m^2}\right)^2 \exp(-D(r_n, |s_i - s_j|)) \\
&= \lim_{m\to\infty} m^2 \sum_{j\in\mathcal{A}_i, j\neq i} \left(\frac{n}{m^2}\right)^2 \exp(-D(r_n, |s_i - s_j|)) \\
&= n \int_{2r_n \leq |x| \leq 5r_n} \exp(-D(r_n, |x|))dx \\
&\leq n\pi(5r_n)^2 \exp\left(-\frac{3}{2}\pi(2r_n)^2\right),
\end{aligned} \tag{3.40}$$

where the last equality follows from the definition of the Riemann integral and the inequality follows from geometry depicted in Figure 3.4. To see that this last expression tends to 0 as $n \to \infty$, substitute $\pi(2r_n)^2 = \log n + \alpha$ twice.

We have shown that both (3.36) and (3.40) tends to zero as $n \to \infty$, hence it follows from Theorem 3.3 that

$$\lim_{n\to\infty} \lim_{m\to\infty} d_{TV}(W_n^m; Po(\lambda)) = 0. \tag{3.41}$$

Since by definition W_n^m converges a.s. to W_n as $m \to \infty$, (3.33) and (3.41) imply that W_n converges in distribution to a Poisson random variable of parameter $e^{-\alpha}$ as $n \to \infty$. □

Having discussed the node-isolation results, we next need to relate these results to the question of full connectivity.

The two results above, given by Propositions 3.10 and 3.11, can be interpreted as the asymptotic almost sure behaviour of the length N_n of the *longest edge* of the nearest neighbour graph of the Poisson points inside B_n. Indeed, the transition from having one to having no isolated points when we let the radii grow clearly takes place when the point with the furthest nearest neighbour finally gets connected.

Now let the *Euclidean minimal spanning tree* (MST) of the Poisson points in B_n be the connected graph with these points as vertices and with minimum total edge length. Let M_n be the length of the longest edge of the MST. The following is a main result in [7].

Theorem 3.12. *It is the case that*

$$\lim_{n\to\infty} P(M_n = N_n) = 1. \tag{3.42}$$

We also have the following geometric proposition.

Proposition 3.13. *If $r_n > M_n/2$, then $G_n(r_n)$ is connected; if $r_n < M_n/2$, then $G_n(r_n)$ is not connected.*

Proof. Let $r_n > M_n/2$. Note that any two points connected by an edge in the MST are within distance $d \leq M_n$. It immediately follows that $MST \subseteq G_n(r_n)$ and hence $G_n(r_n)$ is connected. Now let $r_n < M_n/2$. By removing the longest edge (of length M_n) from the MST we obtain two disjoint vertex sets V_1 and V_2. Any edge joining these two sets must have length $d \geq M_n > 2r_n$, otherwise joining V_1 and V_2 would form a spanning tree shorter than MST, which is impossible. It follows that $G_n(r_n)$ cannot contain any edge joining V_1 and V_2 and it is therefore disconnected. □

To conclude the proof we note that it follows from the last two results that w.h.p., if $r_n > N_n/2$, then the graph is connected, whereas for $r_n < N_n/2$ it is not. But we noted already that $r_n > N_n/2$ means that there are no isolated points, while $r_n < N_n/2$ implies that there are. □

Chapter 4

Simulations

In this last section of the thesis, we have made some simulations to verify the main results that we have discussed in the thesis.

The parameters in the simulations are:

- the radius of the circles r;
- the size of the squared domain $\Lambda = [0, T] \times [0, T]$, with $T > 0$ large;
- the density $\lambda > 0$ of the Poisson process X on the squared domain.

In every simulation, to create the Poisson process we need to remember two key properties:

- the number of points in Λ has distribution $Poi(\lambda T^2)$;
- conditional on $X(\Lambda)$ (number of points in Λ), locations of points $x \in X$ are independent and uniformly distributed on the domain Λ.

In other words, the number of points in the process can be generated through a Poisson function, since $N \sim \text{Poi}(\lambda T^2)$. After having generated it, we have used a random function to generate the $2N$ coordinates of the points in the square domain. Note that the procedure to generate the Poisson process works only because the domain is a square, to simulate the process in different domains, for example in circles, see [8].

We have made several independent simulations for every choice of the parameters, calculating the mean of the values obtained. In this way we should avoid to obtain data that deviates too much from the average results, since in all the simulations we are interested in the average behaviour of the networks.

In particular we have confronted the results of the simulations with the three main theorems present in the thesis:

(1) Theorem 2.4: In a boolean random network of radius r, exists a critical density $0 < \lambda_c < \infty$ such that $\theta(\lambda) = 0$ for $\lambda < \lambda_c$, and $\theta(\lambda) > 0$ for $\lambda > \lambda_c$.

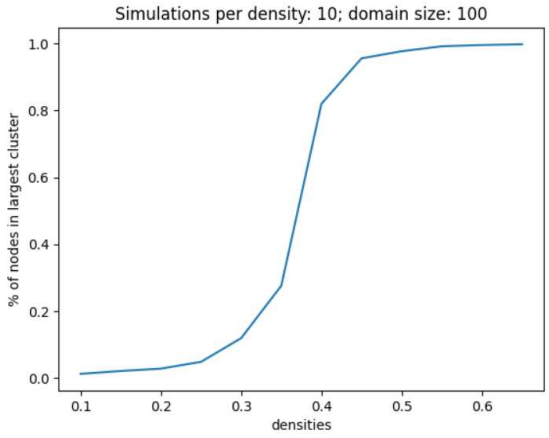

Figure 4.1: Plot showing how the size of the largest cluster in the network increases with the density. ($r = 1$)

(2) Theorem 3.6: We have that for any $\alpha \in (0, 1)$, if $r > r_\alpha$, then $G_n(r)$ is α-almost connected asymptotically a.s., while for $r < r_\alpha$ it is not.

(3) Theorem 3.9: Let $\pi(2r_n)^2 = \log n + \alpha_n$. Then $G_n(r_n)$ is connected w.h.p. if and only if $\alpha_n \to \infty$.

As we can see from the list above, in all the theorems the size of the largest cluster covers a fundamental role. To find it, in the simulations we have created an algorithm inspired by another one that can be found in [4]; see Appendix A.1.

4.1 Phase transition

In this section we want to verify phase transition in the Boolean model. As we know from Theorem 2.4, varying the density or the radius should be indifferent, namely in both cases we should have phase transition. This means that in an infinite Boolean network under a critical density (or critical radius) it is impossible to have an infinite component, while above the critical density (or critical radius) it becomes possible.

In every simulation we have fixed the size of the square domain $T = 100$ and we have found the percentage of points that are part of the largest cluster.

First of all we analyse the behaviour of the network varying the density. In this case we have fixed the radius $r = 1$ and we have simulated densities λ between 0.1 and 0.6, using the program in Appendix A.3. For every density we have made 10 simulations and we have found the average percentage of nodes in the largest cluster.

The results are reported in Figure 4.1.

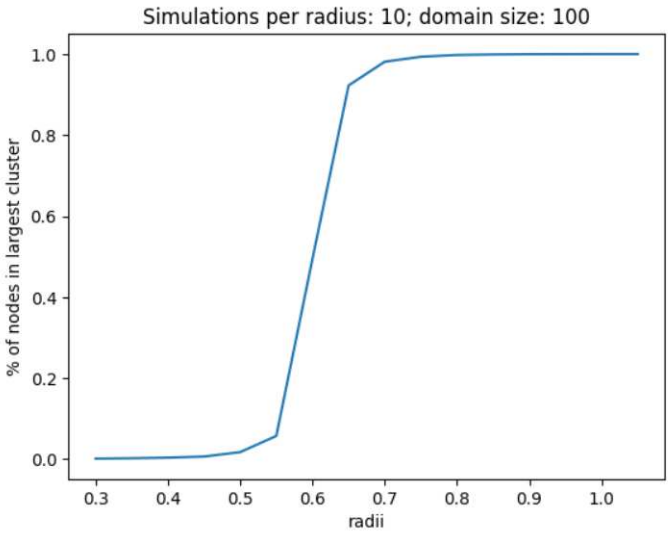

Figure 4.2: Plot showing how the size of the largest cluster in the network increases with the radius. ($\lambda = 1$)

The figure confirms the predictions of Theorem 2.4. It resembles a sigmoid with a very rapid growth, showing that there is a range of values for which we have phase transition. For $\lambda = 0.1$ the percentage of nodes in the largest cluster is approximatively zero, meaning that the network is formed by singletons and isolated clusters with a small dimension. For $\lambda \approx 0.35$ there seems to be phase transition and it is exactly the theoretical value of λ_c predicted from the theorem. Above that value, on the contrary almost all the nodes in the cluster are part of a single connected component. To have even more precise results and to see an even faster growth of the percentage of the nodes in the largest cluster, we could increase the size of the networks.

Lets now analyse the behaviour of the network changing the radius of the circles associated to the nodes of the Poisson process. We have fixed the density $\lambda = 1$ and we have simulated networks with radii between 0.3 and 1, using the program in Appendix A.4. As before, for every value of the radius we have made 10 simulations and we have reported the results in Figure 4.2.

The figure confirms the predictions of Theorem 2.4 for the same reasons as the previous one. It resembles a sigmoid where the percentage of nodes in the largest cluster grows rapidly. The critical value of the radius corresponds to the one predicted from Theorem 2.4: $r_c \approx \sqrt{\frac{4.512}{4\pi}} = 0.6$ Moreover, the results also confirm the predictions of Theorem 3.6, which states that there is a critical value of the radius for which there is α-almost connectivity a.a.s. The fact that the theorem works a.a.s. as the size of the domain tends to infinity implies that to obtain even more precise results we can, for example, increase T, which, however, would also increase the computational time considerably.

4.2 Connectivity

In this section we are interested in the full connectivity of the networks. The question is when all nodes are part of a single cluster w.h.p.

Theorem 3.9 states that there is full connectivity w.h.p. if and only if

$$r_n = \sqrt{\frac{\log n + \alpha_n}{4\pi}}, \tag{4.1}$$

where $\alpha_n \to \infty$ for $n \to \infty$.

To verify this statement we have fixed the density $\lambda = 1$ and we have taken increasing values of the domain size T. For each value of $n = \sqrt{T}$ we can predict from (4.1) the value of r_n for which we should have full connectivity w.h.p. as $n \to \infty$. For every case we have made 10 simulations and we have found the average percentage of nodes in the largest cluster; see Appendix A.5 .

We have simulated three different series of radii

- *Series* 1: $r_n := \sqrt{\frac{\log n}{4\pi}}$, it should be asymptotically under the critical increasing radius since $\alpha_n = 0$;
- *Series* 2: $r_n := \sqrt{\frac{2\log(n)}{4\pi}}$, it should be asymptotically above the critical increasing radius since $\alpha_n = \log(n) \to \infty$;
- *Series* 3: $r_n := \sqrt{\frac{n^{0.3}}{4\pi}}$, it should be asymptotically above the critical increasing radius since $\alpha_n \to \infty$ and in particular the size of the largest cluster should grow faster than the one in the second series.

The results are reported in Figure 4.3 and they confirm the predictions. The largest cluster is initially relatively small for all the series because the initial size of the networks is low if compared with the density and the size of the radii. The plot clearly shows that in the first series the size of the largest cluster remains negligible compared to the number of nodes in the network, confirming that in this case, for $n \to \infty$, the network is formed by small clusters not connected with each other. Instead, for the second and third series we can see that the size of the largest network increases with the size of the networks, in particular for the last one the largest cluster contains almost all the nodes of the network for $T \geq 120$. For having the same results for the second series we should simulate networks with greater size, but this would exponentially increase the computational running time. Despite this, the pattern is still clear, and the plot confirms the predictions of Theorem 3.9.

We can conclude that the results of the simulations coincide with what we expected from the theorems, which are a useful instrument to predict the behaviour of large clusters in Boolean networks.

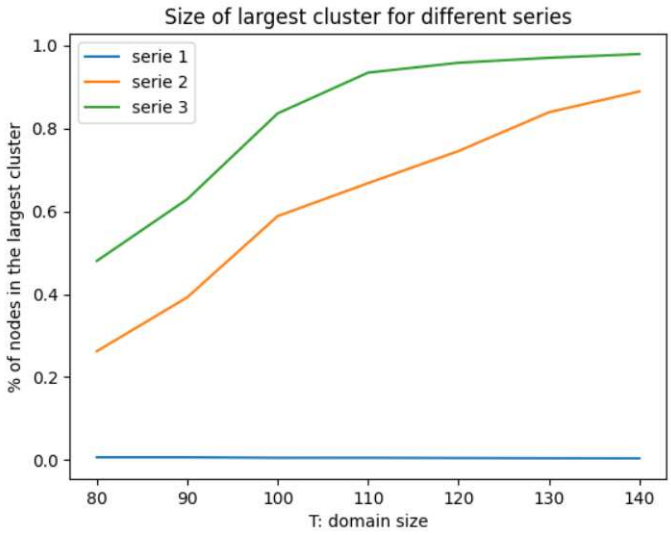

Figure 4.3: Plot showing the variation of the size of the largest cluster in the networks for different series of radii. ($\lambda = 1$)

Appendix A

Algorithms

In this appendix we report the algorithms that we have used in Chapter 4. The algorithms are written using the language Python.

```
#n = number of points in the Poisson process
#coordx, coordy = coordinates of the points
#R = radius of circles

def find_clusters():
  cluster_list = [] #list of clusters
  for i in range(n):
    cluster_i = set() #cluster of i
    for j in range(i,n):
      if (math.dist([coordx[i],coordy[i]],[coordx[j],coordy[j]]) <
    2*R) and (j not in cluster_i):
        for k in cluster_list:
          if j in k:                        #if j is already in a
    cluster, we remove the cluster from cluster_list
            cluster_list.remove(k)
            cluster_i = cluster_i | k     #we do not have to visit
    the other nodes in the cluster
        cluster_i.add(j)
    cluster_list.append(cluster_i)
  return cluster_list
```

Listing A.1: Algorithm for finding clusters

```
# Create limits (x,y)=((-T/2,T/2),(-T/2,T/2))
plt.xlim((-T/2,T/2))
plt.ylim((-T/2,T/2))

# Plot points
for cluster in cluster_list:
  rand_color = "#"+''.join([choice('ABCD0123456789') for i in range
    (2)])*3 #not picking "E" or "F" to avoid colours too bright
  for i in cluster:
    circle = plt.Circle((coordx[i],coordy[i]),R,color=rand_color)
    plt.gca().add_patch(circle)

```

```
plt.show()
```

Listing A.2: Algorithm for plotting the network

```
import numpy as np
import matplotlib.pyplot as plt
import math
from random import *

# Parameters
lambda_list = list(np.arange(0.1,2.5,0.1)) #density of the nodes
num_sim = 10 #number of simulations
T = 100 #domain size
R = 1 #radius of circles
dim_largest_cluster_list = []

for Lambda in lambda_list:
  lam = Lambda * T * T #density of Poisson distribution
  dim_largest_cluster_Lambda = []
  for i in range(num_sim):
    n = np.random.poisson(lam)  #number of points
    coordx = 2 * T/2 *  np.random.random_sample(n) - T/2 # generate random x
    coordy = 2 * T/2 *  np.random.random_sample(n) - T/2 # generate random y
    cluster_list = find_clusters()
    dim_largest_cluster = 0
    for c in cluster_list:
      if len(c) > dim_largest_cluster: dim_largest_cluster = len(c)
    dim_largest_cluster_Lambda.append(dim_largest_cluster/n)
  dim_media = sum(dim_largest_cluster_Lambda)/num_sim
  dim_largest_cluster_list.append(dim_media)

plt.plot(lambda_list, dim_largest_cluster_list)
plt.xlabel('densities')
plt.ylabel('% of nodes in largest cluster')
plt.title('Simulations per density: %d; domain size: %d'%(num_sim,T))
plt.show()
```

Listing A.3: Algorithm for viewing phase transition changing the density

```
import numpy as np
import matplotlib.pyplot as plt
import math
from random import *

# Parameters
Lambda = 1 #density of the nodes
T = 100 #domain size
lam = Lambda * T * T #density of Poisson distribution
radii = list(np.arange(0.6,1.7,0.05)) #radii of circles
num_sim = 10 #number of simulations
dim_largest_cluster_list = []

for R in radii:
  lam = Lambda * T * T #density of Poisson distribution
```

```
  dim_largest_cluster_R = []
  for i in range(num_sim):
    n = np.random.poisson(lam)  #number of points
    coordx = 2 * T/2 *  np.random.random_sample(n) - T/2 # generate random x
    coordy = 2 * T/2 *  np.random.random_sample(n) - T/2 # generate random y
    cluster_list = find_clusters()
    dim_largest_cluster = 0
    for c in cluster_list:
      if len(c) > dim_largest_cluster: dim_largest_cluster = len(c)
    dim_largest_cluster_R.append(dim_largest_cluster/n)
  dim_media = sum(dim_largest_cluster_R)/num_sim
  dim_largest_cluster_list.append(dim_media)

#Plot
plt.plot(radii, dim_largest_cluster_list)
plt.xlabel('radii')
plt.ylabel('% of nodes in largest cluster')
plt.title('Simulations per radius: %d; domain size: %d'%(num_sim,T))
plt.show()
```

Listing A.4: Algorithm for viewing phase transition changing the radius

```
import numpy as np
import matplotlib.pyplot as plt
import math
from random import *

# Parameters
Lambda = 1 #density of the nodes
T = range(80,150,10) #domain size
num_series = 3 #number of series of radius to test
num_sim = 10 #number of simulations

#Different series of radius to test full connectivity
radii = [[] for i in range(num_series)]
cluster_size_list = [[] for i in range(num_series)]

for i in T:
  radii[0].append(math.sqrt(math.log(i)/(8*math.pi))) #under the critical value
  radii[1].append(math.sqrt(math.log(i)/4*math.pi)) #above the critical value
  radii[2].append(math.sqrt(i**0.3)/(4*math.pi))) #above the critical value
  for j in range(num_series):
    large_cluster_list = []
    R = radii[j][-1]
    for k in range(num_sim):
      lam = Lambda * i * i #density of Poisson distribution
      n = np.random.poisson(lam)  #number of points
      coordx = 2 * i/2 *  np.random.random_sample((n)) - i/2 #generate random x
      coordy = 2 * i/2 *  np.random.random_sample((n)) - i/2 #generate random y
      cluster_list = find_clusters()
```

```
      largest_cluster = 0
      for c in cluster_list:
        if len(c) > largest_cluster: largest_cluster = len(c)
      large_cluster_list.append(largest_cluster/n)
    cluster_size = sum(large_cluster_list)/num_sim
    cluster_size_list[j].append(cluster_size)

plt.plot(T, cluster_size_list[0], label = "serie 1")
plt.plot(T, cluster_size_list[1], label = "serie 2")
plt.plot(T, cluster_size_list[2], label = "serie 3")
plt.xlabel('T: domain size')
plt.ylabel('% of nodes in the largest cluster')
plt.title('Size of largest cluster for different series')
plt.legend()
plt.show()
```

Listing A.5: Algorithm to compare network's connectivity for different series

Bibliography

[1] Richard Arratia, Larry Goldstein, and Louis Gordon. 'Two moments suffice for Poisson approximations: the Chen-Stein method'. In: *The Annals of Probability* (1989), pp. 9–25.

[2] François Baccelli and Bartłomiej Błaszczyszyn. 'On a coverage process ranging from the Boolean model to the Poisson–Voronoi tessellation with applications to wireless communications'. In: *Advances in Applied Probability* 33.2 (2001), pp. 293–323.

[3] Massimo Franceschetti and Ronald Meester. *Random networks for communication*. Cambridge University Press, 2007.

[4] Magnus Lie Hetland. *Python Algorithms*. Apress, 2010.

[5] Karl Kurten. 'Correspondence between neural threshold networks and Kauffman Boolean cellular automata'. In: *Journal of Physics A: Mathematical and General* 21.11 (1988), p. L615.

[6] Ronald Meester and Rahul Roy. *Continuum percolation*. Vol. 119. Cambridge University Press, 1996.

[7] Mathew Penrose. 'The longest edge of the random minimal spanning tree'. In: *The annals of applied probability* 7.2 (1997), pp. 340–361.

[8] Sheldon Ross. *Simulation*. Academic Press, Inc., 2006.

[9] David Vere-Jones and Daryl Daley. *An introduction to the theory of point processes*. Springer, 1988.

[10] Rui-Sheng Wang, Assieh Saadatpour, and Reka Albert. 'Boolean modeling in systems biology: an overview of methodology and applications'. In: *Physical biology* 9.5 (2012), p. 055001.