An Analysis of Spatial Percolation Structures Using a Network Approach

by

Mohammed Altaj Mohammed Fadul

Thesis presented for the degree of Masters of Science

Department of Mathematics and applied Mathematics

University of Cape Town

July 2006

Declaration

I know the meaning of plagiarism and declare that all the work in the document, save for that which is properly acknowledged, is my own.

Signed..... Date.....

M. A. Mohammed Fadul

Abstract

In this thesis we analyse several spatial structures, built from percolation models, by means of an approach used so far in the field of network science. In the first chapter we summarize the major network concepts and characterizations that have been obtained as regards the statistical properties of several data sets or theoretical models. We also give a brief introduction to percolation theory and its applications, adding details in two particular cases where mathematical results are available. In the second chapter we then study one particular application of percolation theory to the modelling of distribution and species abundance at different scales. We mainly focus on the way percolation theory was used to compare two different spatial patterns, particularly the random and the aggregated distribution. Our contribution consists here in showing that percolation theory cannot be used to discriminate between these two patterns, without carrying out further studies explicitly introducing aggregation in the percolation model. In chapter 3, we further study this classical site percolation model defined on a 2D square lattice, but we focus here on its statistical properties as a Spatial Network (SN). We compare these statistical properties (degree distribution, cluster size distribution and occurrence of a giant cluster, shortest path length distribution) to those of the so-called random network (RN) whose properties have been already extensively studied and are revisited here. In chapter 4, we introduce two new types of networks built from a multiscale percolation structure. The first type of construction leads to a spatial network accounting for spatial aggregation. The second type is associated to the multiscale percolation structure in a way which

accounts for the hierarchical occurrence of an increasing number of nodes associated to areas of decreasing sizes and decreasing number of neighbours. In all the studied cases, hints coming from exploratory computer simulations and some attempts of mathematical calculations confirm that spatial networks, either random, aggregated or multiscale may exhibit very different statistical properties than those obtained so far on other types of classical networks. The results presented here in some simple case studies may later provide interesting approaches to model some spatial networks from the real world.

Dedication

To my, dear parents,

brothers, sisters

and

friends.

Acknowledgements

I would like to express my gratitude to my supervisors, Dr. G. Witten and Dr. E. Perrier, whose support, stimulating suggestions and ecouragment made this experience all the more worthwhile.

I am deeply indebted to my co-supervisor Dr. E. Perrier. I have known her as a principle-centered person. Her integral view on research and her mission for providing, only high quality work, has made a deep impression on me. She could not even realise how much I have learned from her, I am really glad that I have come to get know Edith.

I would like to acknowledge the financial support provided by the African Institute for Mathematical Sciences (AIMS) South Africa, towards my Masters studies.

My acknowledgement also goes to all my friends, whose gave me the possibility to complete this thesis, thank you for the unity and being a source of inspiration to me.

To you all words really fail me on how to thank you. However, may Allah continue blessing you abundantly.

Contents

List of Figures

List of Tables

CHAPTER ONE

Network science and percolation theory: A literature review

This chapter presents a literature review of network science and percolation theory and their application. The outline of this chapter is as follows: First, we will introduce general definitions and notations in the field of networks, history of networks, why this subject has become an interesting field for research, and give some examples of certain types of real-networks. We will then discuss some of the statistical properties used to describe network topology, such as path length, clustering coefficient, degree distribution, network resilience, degree correlation and the giant component, that characterize the structure and behavior of networked systems [1]. Also, we will give the general types of the networks and their properties. Then, we will give a brief definition of percolation theory and its applications. Here we will consider two particular cases; percolation in one dimension and the Bethe lattice. We shall then link them to the network science.

1.1 Introduction to networks

In mathematical terms [4] a network may be represented as a graph that consists of two sets: a set of nodes (vertices, or points), $V = (v_1...v_N)$, which is the fundamental unit, and a set of edges (links), $E = (e_1...e_E)$, between pairs of nodes. There are N nodes and E edges, where each edge can be written as a pair of nodes, $e_k = (v_i, v_j)$. Two nodes joined by an edge are said to be adjacent or neighbours. The way the nodes and links are located are, in most cases, irrelevant. The only thing that matters is which pairs of nodes are connected to form a link and which are not [4]. However, this is not the case in some spatial networks. In this case, the location of the nodes is important and the type of connection between them depends on the actual neighbouring or geographical space (as we have described in Chapter 3).

A graph is said to be connected when there is at least a link between a node to any other node in the graph. A graph that is not connected is said to be disconnected. It is said to be directed or (digraphs) if its pointing in only one direction. In this case, the order of the two nodes in an edge is important. The graph is said to be undirected if it runs in two directions *i.e* in graphs where each edge consists of an unordered pair of nodes, implying that the edge between v_i and v_j is identical to the edge between v_j and v_i. A weighted graph is that which has a numeric label $w(e)$ associated with each edge. This weight can be a distance, connection cost (*Fig. 1.1*). A graph is said to be simple if it has at most one edge between each pair of nodes and no edge connects a node to itself (loop). Furthermore, a graph is referred to as random if some kind of randomness

is involved in its construction [1].

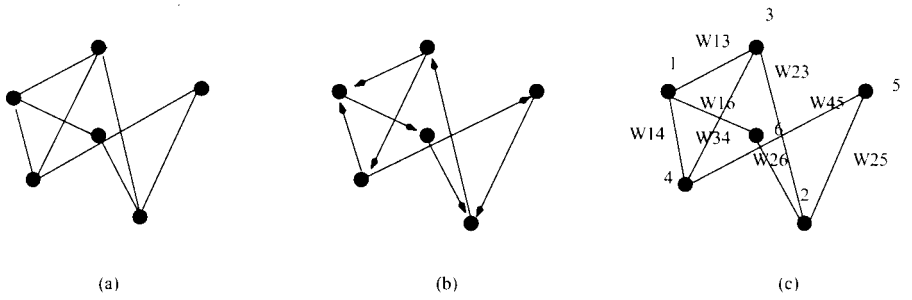


Figure 1.1: *Graphical representation of an undirected (a), a directed (b), and a weighted undirected (c) graph with $N = 6$ nodes.*

Recently there has been a great interest in the field of networks. The conception of the theory is universally attributed to Euler and his solution of the celebrated Königsberg bridge puzzle [5]. Erdös interest on network theory deals with general communication networks. This problem was formalized by Kochen and Pool in the 50's, leading them to the definition of random graphs [5] in which the existence of a link between any pair of nodes has probability p. Erdös, in collaboration with Rènyi, pursued the theoretical analysis of the properties of random graphs and obtained a number of important results [5]. Random graph theory has since produced a prodigious number of results, describing statistical properties of graphs, such as distribution of component (cluster) sizes, existence and size of a giant component [6], path length and degree distributions. There exist a number of reviews of the area of network science. Albert and Barabasi [4] and Dorogovtsev and Mendes [7, 8] have given good reviews focusing on the physics literature and models of growing networks. Other reviews focusing on one type of network has been given by

Newman [9] and Hayes [10, 11] who concentrate on the small world models. A good and interesting discussion of the behavior of the dynamical systems on network have been given by Strogatz [12], dynamical systems can often be mofelled by differential equations $\frac{dx}{dt} = v(x)$. where $x(t) = (x_1(t), x_2(t), ..., x_n(t))$ is a vector of state variables, t is time, and $v(x) = (v_1(x), v_2(x), ... v_n(x))$ is a vector of functions that encode the dynamics. These diferential equations are oftenly nonlinear. We will focus in the present work on static networks.

Recent work on the mathematics of the networks has been driven largely by observations of the properties of actual networks, ranging from communication networks to ecological webs and attempts to model them. Table 1.1 represents examples of some kinds of real networks.

In general, there are two types of networks, static networks and complex dynamical networks. In static networks the number of nodes are fixed. This type of network is a good model for cases in which growth or aging process do not play a dominant role in determining the structural properties of the network [27]. Examples of such cases will be considered in Chapters 3 and 4. Complex dynamical networks represent all kinds of real-world networks such as biological, physical, neural networks, the internet and the WWW. All these are examples composed of a large number of interacting dynamical units that together produce global complex systems. The main goal of studying these networks of interacting dynamical systems is to understand the global behavior of the network and the extent to which the dynamics of the individual nodes and the network structure are responsible for the global behavior. In the following section we will describe some of the

Network	**Type**	**Nodes**	**Edges**	**Degree**	**Ref**
WWW	directed	webpages	hyperlinks	power-law	[1, 4, 13, 14] [15, 16, 17, 18]
Internet	undirected	routers	physical link	power-law	[1, 4, 19]
Ecological network	undirected	species	predator-prey	power-law	[1]
Cellular network	directed	substrates	chemical reaction	power-law	[4, 20]
Power grid	undirected	generators ,tansformer	high-voltage transmission	-	[21]
Movie actors	undirected	actors	actors have acted in a movie together	power-law	[22, 23]
Sexual contact network	undirected	human person	physical relations between persons	power-law	[24, 25, 26. 1]

Table 1.1: *Different examples of real world networks.*

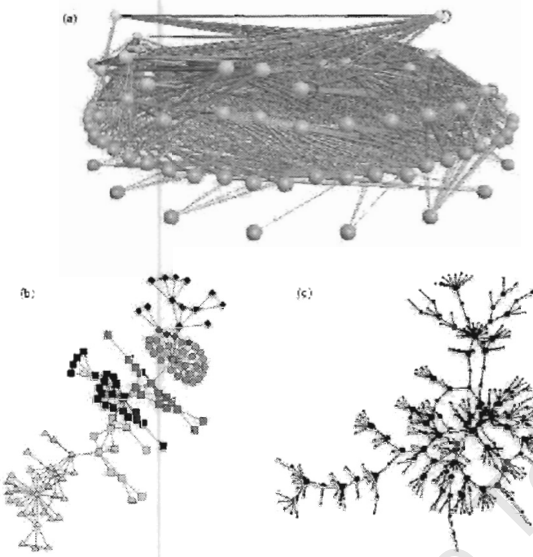


Figure 1.2: *This figure is extracted from the excellent Newman review [1]. It shows three examples of the kind of networks mentioned in table 1.1. (a) A food web of prey-predator interactions between species. (b) The network of collaborations between scientists at a private research institution. (c) A network of sexual contacts between individuals.*

statistical properties of networks.

1.2 Some statistical properties of networks

There are many measures for networks that have been proposed and investigated. In this section we describe measures that could be common to networks of many different types. By measuring these properties one can understand the structure of a network. We shall define and briefly discuss them. Further discussions will follow in later sections of this thesis.

1.2.1 Path length

The characteristic path length L is defined as the *average distance between any two nodes*, or more precisely, *the average length of the shortest path connecting each pair of nodes over all pairs of nodes in the network*. For undirected and unweighted network, the distance between two nodes is simply the minimal number of edge. that are required to go from one node to the other [1]. For the unconnected nodes, the path length said to be unbounded. The distance between the two nodes v_i and v_j is denoted by l_{ij}. The characteristic path length L is generally used to measure the global properties of a graph and is independent of local structure:

$$L = \frac{1}{\frac{N(N-1)}{2}} \sum_{v_i, v_j \in V} l_{ij},$$

where l_{ij} = min (all the length of the all possible paths going from nodes i to j).

1.2.2 Clustering coefficient

The clustering coefficient for a node v_i is the proportion of links between the nodes within its neighbourhood divided by the number of links that could possibly exist between them [28]. It quantifies how well connected the neighbours of a given node in a network are [29] and describes the cluster and the local structure of the graph. Focusing on a selected node v_i in the network with k_i neighbours, there would be, in the fully connected network, $k_i(k_i - 1)/2$ edges between the nodes within the neighbourhood. Thus the clustering coefficient is given as [4]

$$C_i = \frac{2E_i}{k_i(k_i - 1)},$$

where E_i is the number of edges that actually exist in the subnetwork made by the closest neighbours of node i. The clustering coefficient of the whole network, is then the average of this interconnection of node neighbours, and is equal to

$$C_{WS} = \frac{1}{N}\sum_{i=1}^{N} C_i. \tag{1.1}$$

The subscript $_{WS}$ is included to distinguish the *Watts-Strogatz* clustering coefficient.

There is an alternative definition for clustering coefficient for a whole network as:

$$C = \frac{3(\text{number of triangles in the graphs})}{(\text{number of connected triples})}. \tag{1.2}$$

A triangle is a set of three nodes (v_j, v_k, v_l), where each node is connected to the other two. The denominator determines the number of connected triples, where j is connected to k, and k is connected to l, although l does not necessary have to be connected to j.

$$C = \frac{3\sum_{i,j\neq 0} p_{0i}p_{0j}p_{ij}}{\sum_{k,l\neq 0} p_{0k}p_{0l}} \tag{1.3}$$

The denominator gives the expected number of connected triples, while the numerator determines the expected number of triangles. The factor 3 in the numerator is necessary because each triangle gives rise to 3 connected triples. When $C_{WS} = 1$, it means that the network is fully connected, and when $C_i = 0$ for some i, the network is disconnected.

1.2.3 Giant component (cluster size)

A cluster is a subset of the network where there exists at least a path between any two nodes, that is, a group of connected nodes. The cluster size is the number of nodes in this group, in some networks the study of cluster size distribution have been widely studied.

Particularly the size of the largest cluster is an important quantity. The behaviour of the cluster size distribution above and below a critical density of links at which a giant cluster first forms is also important(see chapter 3).

1.2.4 Degree distribution

Information of a network can be also extracted from its degree distribution, because often this statistical measure gives us a good sense of what the graph looks like. Not all the nodes in the network have the same number of links, so there is a spread in the number of links associated with a given node. For any given network we can measure the degree distribution by making a list of the degrees (the number of links attached to a given node) that occur in the graph and the frequencies with which they appear or, in other words, the probabilities for each node to have a degree k. We define p_k to be the fraction of vertices in the network that have degree k. The degree distribution of a network also has a profound effect on the way the network behaves. In particular, the existence of nodes in a network which has a very large number of connections can wholly alter the character of the network's dynamic. This effect has been observed, for example, in the behaviour of WWW, the spread of sexually transmitted diseases and in food webs [30].

1.2.5 Degree correlation

In most kinds of networks there are different types of nodes, and the probability of having two nodes connected depends on the types. This kind of selective linking is called assortative mixing [1]. For example, nodes with high degree may have a larger probability to be

connected with large degree (social networks) [1, 31]. An example of assortative mixing is social networks with mixing by race. A special case of assortative mixing according to a scalar node property is mixing according to node degree. There are two situations that have been shown, assortative mixing and disassortative mixing. the assortative mixing by degree is of particular interest because. since degree itself a property of the network topology. degree correlations can give rise to some interesting netowrk structure effects. Disassortative mixing refers to a situation where high degree nodes have majority of neighbours with low degree (information networks, biological networks. technological networks) [1. 32].

1.2.6 Network resilience

The resilience of networks to the removal of their nodes has been studied in a wide way in the literature. One can look for the importance of this property by considering the existence of paths between two given nodes v_1 and v_2 in the networks. If other nodes are removed from the network, the length of paths between v_1 and v_2 may increase, and by continuously removing nodes, eventually the two selected nodes v_1 and v_2 become disconnected. Not all the networks have the same level of resilience to such nodes removal. Nodes in the networks can be removed in different ways. For example, by random nodes removal. or by targeting some specific class of nodes such as nodes with highest degree. Most networks are robust against random nodes removal. and less robust to targeted removal [1]. Many authors have looked at specific networks and studied the effects of this property such as in the internet network [33]. metabolic networks [20]. food webs [34, 35]

and email networks [36].

1.3 Networks models

More attention has been paid to some types of networks than others because they can be linked to real-world networks. Next, we will discuss the most famous models of networks and their properties.

1.3.1 Random networks and their properties

The first serious attempt at constructing a model for large random networks was the random net of Rapoport and collaborators [37, 38, 1], which was independently rediscovered by Erdös and Rényi [39]. They defined a random graph as N connected nodes with n edges, which are chosen randomly from the $\frac{N(N-1)}{2}$ possible edges [4]. In total, there are $C^{n}_{\frac{N(N-1)}{2}}$ different graphs with N nodes and n edges. The term random graph refers to the disordered nature of the arrangement of links between different nodes. To avoid confusion

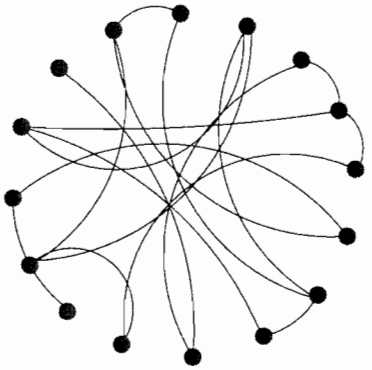


Figure 1.3: *Random Network*

with other types of network such a random graph is alternatively called a binomial model [1]. where some randomness also occurs, but because this example is widely used and extensively studied it is referred to as the "random network" and we will use this word throughout the present thesis. We start with N nodes, and connect each pair of the nodes with probability p. If G is a graph with N nodes and n edges, the probability of obtaining it by this graph construction process is binomial [4] :

$$P(G) = p^n(1-p)^{\frac{N(N-1)-n}{2}}.$$

In the mathematical literature, the construction of a random graph is often called the evolution process. Starting with a set of N vertices, the graph develops by the successive addition of random edges. The graphs obtained at different stages of this process correspond to larger and larger connection probabilities p, eventually obtaining a fully connected graph (having the maximum number of edges $\frac{N(N-1)}{2}$ as $p \to 1$).

The properties of such graphs have been studied extensively because of their great mathematical interest. Random networks are poorly clustered. As Watts and Strogatz [21] showed, random graphs have clustering coefficient $C_{WS} \sim \frac{k}{N} \ll 1$, and have average path length which is approximately given by $L \sim \frac{ln(N)}{ln(k)}$, where k is the number of edges per node (the total number of neighbours for specific node).

In large random graphs, there are several nodes with the same degree, and the number of nodes with a given degree (degree distribution) can be calculated. Erdös and Rényi were the first to study the distribution of the maximum and minimum degree in random graph [39], followed by Bollobás [40] who gave the full degree distribution. The degree

distribution for an infinite random graph is Poissonian,

$$P(k) = \frac{e^{-\lambda}\lambda^{k}}{k!}.$$

where λ is the mean degree. For this reason, ER graphs are also sometimes called Poisson random graphs. ER random graphs are uncorrelated graphs, the edges are connected to nodes regardless of their degree. Random graphs are inadequate to describe some important properties of real world network, so they have been extended in a variety of ways.

One of the most interesting findings of random graph theory is the existence of a critical probability at which a giant cluster forms. Translated into network language, it indicates the existence of a critical probability p_c in a random network such that below p_c, the network is composed of isolated clusters, but above p_c, a giant cluster spans the entire network, which means, most of the nodes are connected to each other. This phenomenon is markedly similar to a percolation transition, a topic much studied both in mathematics and in statistical mechanics [4]. In *Scc.1.4* we will give a general definition of percolation theory and it is application.

1.3.2 Regular networks and their properties

A regular network is defined as a graph where each node has exactly the same number of neighbours. Examples of regular networks are rectangular grids; where all nodes have exactly four links and hexagonal lattices, where each node is connected to three other nodes. Passing from one node to a distant one through intermediate nodes makes long

path lengths. Regular networks are highly clustered, *i.e.* there is a high density of connections between nearby nodes. Watts and Strogatz [21] built a one dimensional (See *Fig. 1.4*) with a finite number of nodes regularly distributed on a ring. They showed that this network has a clustering coefficient $C \sim \frac{3}{4}$, and a path length $L \sim \frac{N}{2k} \gg 1$, where k is the number of edges per vertex. The degree distribution of regular networks is the uniform distribution.

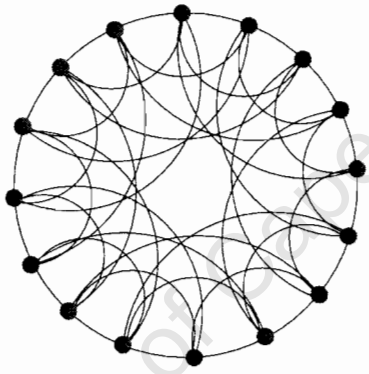


Figure 1.4: *Regular Network*

1.3.3 Small-World networks and their properties

Small-world network means that every node can be reached from every other node by a small number of steps (links) or, more generally. a small-world network is a graph with a combination of a short characteristic path length and generally high degree of clustering.

Although regular networks and random graphs are both useful idealisations, many real networks lie somewhere between the extremes of order and randomness. To interpolate between regular and random networks, Watts and Strogatz [21] studied what happen by

modifying an initial regular network. They considered the following rewiring procedure:

- Start from a ring lattice with k edges per vertex.
- Rewire each edge at random with probability p.

Thus, if $p = 0$, the original regular network is unchanged, but if $p = 1$, the resulting network is completely random. For $0 < p < 1$ they can build small-world network. The small-world property in real networks is associated with a high clustering coefficient, like regular lattice, but short path lengths between nodes. So, transitioning from a regular lattice to a small world topology strongly affects the properties of the graph. For example, a small fraction of random links added to a regular lattice allows disease to spread much more rapidly across the graph [21].

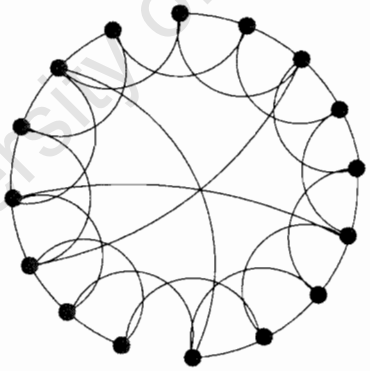


Figure 1.5: *Small-World Network*

Intuitively, one can see why small-world networks might provide a good model for a number of situations. For example, people tend to form tight clusters of friends and colleagues (a regular network), but then one person may move from his city to another city, introducing a random edge.

The results of Watts and Strogatz then provide an example for the empirically observed phenomenon that there often seems to be surprisingly short connections between unrelated people. Many empirical examples of small-world networks have been documented in fields ranging from cell biology to business. On the theoretical side, small-world networks are turning out to be interesting to different scientists. For example, epidemiologists have asked how local clustering and global contacts together influence the spread of infectious disease, with implications for vaccination strategies and the evolution of virulence [12]. The neural network of a nematode worm, and the internet movie database all have the same two characteristics.

The degree distribution for the small world network, in the original Strogatz-Watts-Newman model [30, 21, 41] , with $k = 1$, meaning that each node is connected to its two neighbours before rewiring the nodes, is given by

$$P(k) = (1 - p)f_k(p) + pf_{k-1}(p).$$

where

$$f_k(p) = \frac{f_k(p)e^{-p}p^k}{k!}$$

is the Poisson distribution.

Instead of rewiring each edge with a probability p, Newman and Watts derived a slightly different procedure for making a small world graph, which is simply by inserting a shortcut between any two nodes that have been chosen randomly [21].

1.3.4 Scale-Free networks and their properties

In some real networks, whereas some nodes have a tremendous number of connections to other nodes, most nodes have just a handful. The popular nodes, called hubs, can have hundreds, thousands or even millions of links. In this sense, the network appears to have no scale [42]. So, scale free networks are characterised by an uneven distribution of connectedness. Instead of the nodes of these networks having a homogeneous distribution, some few nodes act as very connected hubs.

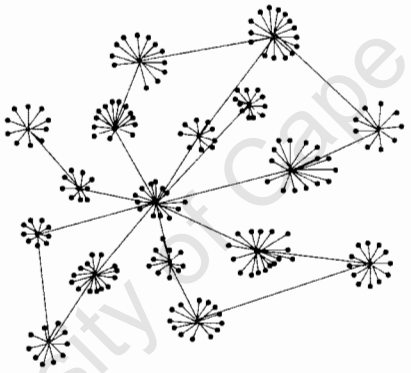


Figure 1.6: *Scale-free Network*

In the power-law distribution, which means, most of the vertices have the same low degree but a small number of vertices have a much higher degree. It was found that most of the real world networks display such a power-law degree distribution $P(k) = Ak^{-\gamma}$, where γ is the degree exponent. The value of γ determines many properties of the system. The smaller the value of γ, the more important the role of the hubs is in the network. Whereas $\gamma > 3$ the hubs are not relevant, for $2 < \gamma < 3$ there is hierarchy of hubs, with the most connected being in contanct with a large fraction of all nodes. Power-laws have

a particular role in statistical physics because of their connections to phase transitions and fractals [27]. The earliest and most basic model generating scale-free network was introduced by Barabaási and Albert [42] by considering two main ingredients, growth and preferential attachment. The algorithm used is as follows :

- Start from small number m_0 of disconnected nodes.
- Add a new vertex every step with m links that that are connected to an old node.
- When choosing the nodes to which the new node connects, we assume that the probability P that a new node will be connected to node i depends on the degree k_i of node i, such that $P \sim \frac{k_i}{\sum_i k_i}$.

The probability that a node obtains an additional link is proportional to its current degree [43]. Real world networks that are best modeled by scale-free structure include the WWW [42], the physical structure of the internet [14], social networks [42], electrical power grids [44], stock market [45], cancerous cells and the dispersal of sexually transmitted disease [25].

There are two different kinds of scale-free networks :

- Scale-free networks with no local clustering produced by the Barabási Albert model.
- Scale-free networks with high clustering properties as in the model by Klemm and Eguiluz [43].

Scale-free networks have certain important characteristics. They are, for instance, robust against accidental failures (removal of randomly selected nodes) but vulnerable to

coordinated attacks (the removal of few key nodes in the network) [33]. Understanding such characteristics could lead to new applications in many areas. For example, computer scientists might be able to devise more effective strategies for preventing computer viruses from crippling a network such as the internet [42].

Random, small-world and scale free networks have been mainly studied by mathematician, computer scientists, biologists or social scientists. We now move to another type of network mainly studied by physicists and also used in a broad way of application, that is percolation networks.

1.4 Introduction to percolation theory

Percolation theory is one of the simplest models exhibiting a phase transition, and the occurrence of critical phenomena is central to the appeal of percolation. It has been used to model the spreading of oil in water [46], to estimate whether one can build nondefective integrated circuits [47], to model the spread of infections [48] and forest fires [46]. From a mathematical point of view, percolation theory is attractive because it exhibits relations between probabilistic and algebraic (topological) properties of graphs. A good introduction to percolation theory can be found in [49, 50, 51], a short review is in [52]. In the following sections, we will give a brief definition of percolation theory and its application. We will discuss two cases of percolation theory: the percolation in one dimension, because it somehow has relation with the regular networks, and the Bethe lattice percolation because it is related to infinite networks.

Percolation theory describes the behavior of connected clusters in various types of more or less connected regular lattice and in a random graph [53]. It deals with the effects of varying the richness of interconnections present in a random system [54]. Let us assume we have some porous material and we pour some liquid on top. Will the liquid be able to make its way from hole to hole and reach the bottom ? What is the probability that an open path exists from the top to the bottom? The remarkable thing is that many results can often be encapsulated in a small number of simple algebraic relationships. There are many different variants which turn out to be identical in almost all important aspects.

The best way to learn about percolation theory is by example. Imagine that we have a large array of squares. And let us randomly fill in some of the squares (occupied) whereas the other are left empty, as in *Fig. 1.7.*

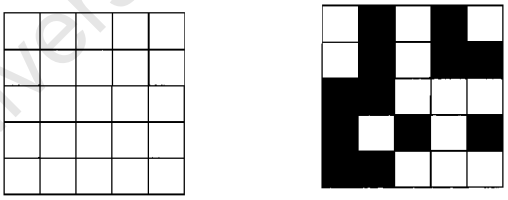


Figure 1.7: *Illustration of site percolation in 2D square lattice with occupied black and empty white site. Each site is occupied with probability p.*

Let each site in the lattice be occupied at random with probability p. That is, each site. occupied (with probability p) or empty (with probability $(1-p)$). is independent of the status (empty or occupied) of any other site in the lattice. We call this p *the*

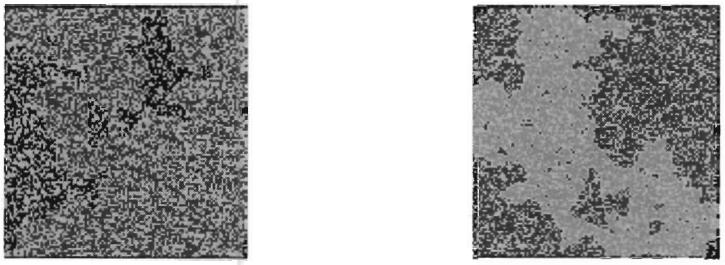


Figure 1.8: *This figure is extracted from [2], it shows the configurations of the percolation problem. The left figure shows a configuration generated with occupation probability* $p = 0.5$. *The right figure shows a configuration generated with occupation probability* $p = 0.6$. *The largest cluster is marked.*

occupation probability or *concentration*. For small values of this probability, we see mostly isolated occupied sites with occasional pairs of neighbouring sites that are both occupied. If neighbouring sites are both occupied, we call it *a cluster*. We can explain more precisely what we mean by clusters by saying: A cluster is a group of occupied sites connected by nearest neighbours distances. As the occupation probability (concentration) increases, we get more clusters. So the cluster on the whole gets larger, see *Fig. 1.8*. If the probability p of the site being occupied increases, several things happen. Most importantly, at a particular value of occupation probability, many of the sites become joined into one giant cluster that spans the entire array both vertically and horizontally [55]. This critical value of occupation probability is known as *percolation threshold* (p_c). Percolation theory deals with the number and properties of the clusters and the value of percolation threshold, and studies the topology of the lattice in the vicinity of the percolation threshold.

the conditions that are used to decide whether a tree ignites its neighbour, one can account for environmental influences such as the direction of the wind or the humidity of air.

- Oil fields: Percolation theory has also been applied in the oil industry to predict the amount of oil that a well will produce. Percolation theory can be used in this case as a simple model for the distribution of the oil or gas inside porous rocks or oil reservoirs. The probability that a site is occupied in the percolation problem corresponds to the porosity or the average concentration of oil in the rock. In order to obtain good oil production from a well, it is desirable to position the well in an area with high porosity. In order to predict the amount of oil that will be produced, one needs to estimate the porosity of the rock in the area where the oil reservoir is assumed to be located.

- Percolation theory in Physics and Chemistry: In order to present the application of percolation theory to physics or chemistry, we are going to start by considering a phenomenon which almost everyone is acquainted with and which can be approached using percolation theory. This phenomenon is the boiling of an egg or, more precisely, the process through which the yolk and the albumen change their consistency. The considered phenomenon is in fact gelation, the formation of a network of chemical bonds, which span the whole system. The interest in such problems started with the study of polymerization. The initial attempts tried to understand how small branching molecules form larger and larger macromolecules

if more and more chemical molecules are formed between the original molecules. In this situation, the initial small molecules correspond to the sites of the lattice and the macromolecules to the clusters.

1.5 The mathematics of percolation

1.5.1 Definition and example

The literature devoted to percolation theory is enormous, and spans the fields of mathematics and statistical physics, as well as a variety of engineering and other applied disciplines [64]. As we mentioned before, percolation theory studies the emergence of paths that percolate through the lattice, for small p only a few edges are present, thus only small clusters of nodes connected by edges can form, but at the critical probability p_c, a percolating clusters appears. This clusters is also called an infinite clusters, because its size diverges as the size of the lattice increases [4]. In the following section we will discuss the central quantities of interest in percolation theory.

1. The percolation probability, P. From the fact that not all occupied sites are in the infinite cluster, we may define the *strength of the infinite cluster* or *Percolation probability* as the probability that an arbitrary site belongs to the percolation cluster (infinite cluster), ($P(p)$ the ratio between the size of the largest cluster and the lattice size)

$$P = \frac{\text{number of sites in the spanning cluster}}{\text{total number of occupied sites}},$$

or

$$P = P_p(|C| = \infty) = 1 - \sum_{s<\infty} P_p(|C| = s)\,. \tag{1.4}$$

where $P_p(|C| = s)$ denotes the probability that the cluster at the origin has size s.

It is clear that below percolation threshold $p < p_c$, this probability $P(p)$ must be zero since there is no infinite cluster. On the the other hand, for $p > p_c$, the cluster at the origin is infinite :

$$P = \begin{cases} 0 & \text{for} \quad p < p_c \\ > 0 & \text{for} \quad p > p_c \end{cases} \tag{1.5}$$

2. The average cluster size.

3. The cluster size distribution, $n_s(p)$, is defined as the probability of a site being the left hand end of a cluster of size s. Unfortunately, it is not possible to obtain an exact form for the cluster size $n_s(p)$ in $d > 1$ because there is a very large number of different ways in which clusters can arrange themselves, see *Fig. 1.9* below

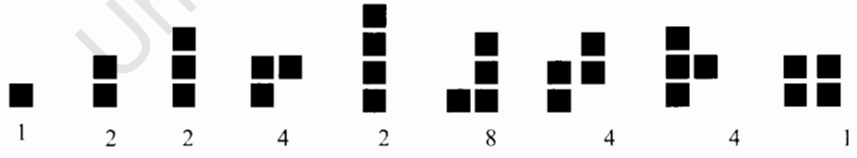


Figure 1.9: *List of all cluster configurations on the square lattice up to* $s = 4$.

Example.

$$For \quad s = 3 :, \quad , \quad n_s(p) = 2p^3(1-p)^8 + 4p^3(1-p)^7$$

To be able to write the general expression for the cluster number, we define *Perimeter (t)* as the number of empty nearest neighbours of the cluster. So

$$n_s(p) = \sum_t g_{s,t} p^s (1-p)^t \tag{1.6}$$

Where $g_{s,t} \equiv$ the number of different lattice configuration with size s and perimeter t.

In one-dimension, all clusters have two perimeter sites and thus

$$g_{s,t} = \begin{cases} 1 & \text{for} \quad t = 2 \\ 0 & \text{for} \quad otherwise. \end{cases} \tag{1.7}$$

implying

$$n_s(p) = p^s (1-p)^2. \tag{1.8}$$

The difficulty of *Eq. 1.6* is that it involves a sum over all possible perimeter t, and thus each possible configuration has to be found and carefully analysed to find the $g_{s,t}$. Tables of such numbers, often in the form of so- called perimeter polynomials,

$$D_s(1-p) = \frac{n_s(p)}{p^s} = \sum_t g_{s,t}(1-p)^t, \tag{1.9}$$

have been published mainly by the King's college group.

These quantities are of interest in random networks as well. There is an important difference between percolation theory and random networks. Percolation theory is defined on a regular d-dimensional lattice. In a random network we can define the neighbourhood of a node as the set of nodes it is connected to, and we can define a non-metric distance along the edges to any other node in the network-there is no regular small - dimensional

lattice a network can be embedded into. However, random network and percolation theory meet exactly in the finite dimensional limit ($d \to \infty$) of percolation [4].

In the following section we will discuss the percolation theory in one dimension and the Bethe lattice, where it can be solved exactly.

1.5.2 Exact solutions for one dimension

For the site percolation problem in one dimension where it can be solved exactly. Some aspects of that solution seem to be valid for higher a dimension also [49]. Imagine a one dimensional lattice with an infinite number of sites of equal spacing arranged in a line. Each of these sites is randomly occupied by probability p. A single empty site would split the group into two different clusters. In order for the cluster to be separated from the other clusters, the site neighbouring the left end of the cluster must be empty, and this is true for the right end of the cluster.

Figure 1.10: *percolation in 1D. Black sites occupied with probability p, empty sites denoted by white, there is one cluster of size 5, one of size 2, two of size 1*

Looking at *Fig. 1.10* containing five occupied sites, we need these five sites occupied and their neighbours empty. The probability of each site being occupied is p since all sites are occupied randomly . So from the fact that all sites are independent, p^2 denotes the probability of two arbitrary sites being occupied, while p^5 denotes five sites being occupied. The probability of one end having an empty neighbour is $1 - p$, and for two

ends is $(1-p)^2$. Therefore the total probability that a fixed lattice site is the left end of five cluster is $p^5(1-p)^2$. In our one dimension case the above consideration for five clusters is easily generalised to

$$n_s(p) = p^s(1-p)^2,$$

where $n_s(p)$ denotes the number of s-clusters per lattice site. Clearly for one dimension, when $p = 1$, all sites of the line are occupied and the whole line constitutes a single cluster. For $p < 1$, there will be some holes in the line. In other words, there is no percolating cluster. Thus the percolation threshold is $p_c = 1$.

Note that every occupied site must belong to one cluster since single occupied sites surrounded by empty neighbours are also clusters of size unity. So the probability that an arbitrary site belongs to any cluster is therefore equal to the probability (p) that is occupied. This is also valid for higher dimensions except that one has to take into account the site in the infinite cluster separately. This leads to the fact that not all occupied sites are in the same cluster. The probability that an arbitrary site belongs to any cluster is equal to the probability p that is occupied [49].

Another interesting quantity is the *correlation function or pair connectivity* $g(r)$, which is defined as the probability that a site at position r from an occupied site belongs to the same cluster. Clearly, $g(r = 0) = 1$, since the site is occupied by definition, and $g(r = 1) = p$, the neighbouring site belongs to the same cluster if and only if it is occupied. For a site at position r to be occupied and to belong to the same cluster, this site and $(r-1)$ intermediate sites must be occupied. Thus $g(r) = p^r$.

1.5.3 Exact solution for the Bethe lattice

The mathematics of percolation theory can be traced back to Flory (1941) and Stockmayer (1943), who were trying to explain the process of gelation in which small molecules adhere to larger and larger ones to eventually form a gel. Because these molecules do not loop back on themselves, but rather branch in definitely, they form a special family of lattices known as Cayley trees, also called Bethe lattice [64]. In this special lattice, each site has z neighbouring sites, such that a branch gives rise to $z-1$ other branches (see *Fig. 1.11*). The One-dimensional case is effectively a Bethe lattice with $z = 2$.

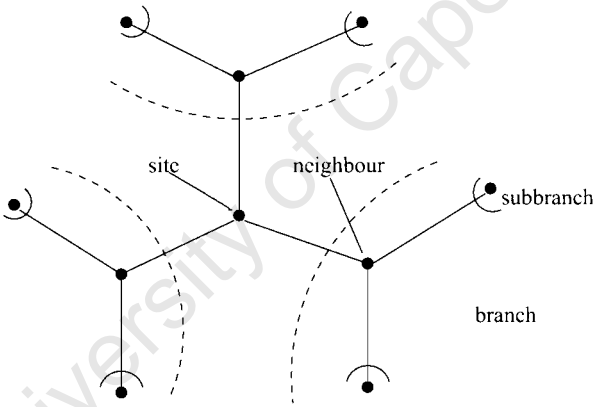


Figure 1.11: *Example of a Cayley tree with neighbouring site $z = 3$. All of the nodes have 3 edges, with the exception of those on the surface, which have only one edge.*

The main point of studying the Bethe lattice is because Bethe lattice solution gives the $d \rightarrow \infty$ limit of percolation, the point where it meets random graphs. For $d-dimensions$, the volume of a sphere is proportional to r^d, and its surface to r^{d-1}. Thus, the surface and volume of a regular d-dimensional object obeys the scaling relation Surface $\propto$ Volume$^{1-\frac{1}{d}}$.

To determine the percolation threshold of the Bethe lattice, we need to start at the origin and check if there is a chance of finding the path of occupied neighbours. Starting from that origin, if we go on such path in the outward direction, one can find $(z-1)$ new neighbours. Thus, on average we have $p(z-1)$ new occupied sites on which we can continue the path. If the number $p(z-1) < 1$, the average number of different paths leading to infinity decreases at each generation by this factor $p(z-1)$. Therefore the percolation threshold is [55]

$$p_c = \frac{1}{z-1} \tag{1.10}$$

Another quantity of interest in studying the Bethe lattice is percolation probability, which is the probability of a site belonging to the infinite cluster. It has been shown that near the percolation threshold, the percolation probability diverges as $p \to p_c$, as follows (the derivation of this equation has been shown in this Ref [55]):

$$P \propto (p - p_c)^{\beta}. \tag{1.11}$$

where β is a critical exponent, and is defined as $\beta = \frac{\tau-2}{\sigma}$.

Mean cluster size is the average number of sites of the cluster to which the origin belongs. Let T be the mean cluster size in one branch (that is the average number of sites to which the origin is connected and which belongs to one branch). Sub-branches have the same mean cluster size as the branch itself. The average cluster size diverges as $p \to p_c$, and it depends on P as a power of the distance $p - p_c$ from the percolation threshold. This behavior is an example of critical phenomena, and the average cluster

What we described above is actually a particular type of percolation theory known as *Site Percolation* on a particular lattice. Table.1.2 shows a list of the percolation threshold in various lattices and dimensions [49]. Instead we could have occupied the edges of the sites (bonds) and this is known as *Bond Percolation.*

The exact value of p_c depends on which kind of grid is used and the dimensionality of the grid. for example the p_c value is not the same in the site and bond percolation. The change of occupation probability values is common in some branches of physics. For example a magnet when heated losses its magnetisation at a particular temperature(the curie temperature). In general this is known as *phase transition or critical phenomena*, which is defined as the transformation of a thermodynamic system from one phase to another. The distinguishing characteristic of a phase transition is an abrupt sudden change in one or more physical properties.

1.4.1 Application of percolation theory

Recently, there has been a great interest in percolation theory and its application to many different fields. Theoretical studies of percolation and the application of percolation models in diverse scientific disciplines have resulted in thousands of papers over the last decade [56]. The principal advantage of percolation theory is that it provides very useful universal laws which determine the geometrical and physical properties of the system [57], and it is one of the simplest models exhibiting a phase transition. Examples of this application include study of percolation of petroleum and gas [58, 59], conductivity [60], polymerization and gelation [61], biological evolution. modeling of the spread of diseases

Table 1.2: *A list of the percolation threshold in various lattices and dimensions*

Lattice	Number of nearest neighbours	Site percolation	Bond percolation
1d	2	1	1
2d Honeycomb	3	0.6962	$1 - 2sin(\frac{\pi}{18}) \approx 0.65271$
2d Square	4	0.592749	$\frac{1}{2}$
2d Triangular	6	$\frac{1}{2}$	$2sin(\frac{\pi}{18}) \approx 0.34729$
3d diamond	4	0.43	0.388
3d Simple cubic	6	0.3116	0.2488
Bethe lattice	z	$\frac{1}{(z-1)}$	$\frac{1}{(z-1)}$

in a population, epidemiology, percolation networks and application of percolation in the field of pharmacy [62]. Most of such applications have resulted in qualitative as well as quantitative predictions for the system of interest. In the following section we describe empirical studies of application of percolation theory [63].

- Forest Fires: The first application of percolation theory in this case refers to the modelling forest fires. The forest is represented by a lattice, whose sites are occupied with probability p by a tree or empty with probability $1 - p$. If $p = 1$, it means that all sites correspond to trees, and the probability $p < 1$ allows for holes (empty sites). A site could be burning, susceptible to burning but not yet ignited ('green'), or burnt out and so not susceptible to ignition. At each moment, a burning site has a fixed probability of igniting a randomly chosen green neighbour. By varying

size is given by [4]

$$S \propto (p - p_c)^{-\gamma}. \tag{1.12}$$

where the exponent γ is defined as $\gamma = \frac{3-\tau}{\sigma}$.

Having earlier defined the cluster size distribution as the probability of a site being the left hand end of a cluster of size s, the general form for the cluster size distribution in the Bethe lattice is expressed as

$$n_s(p) \propto s^{-\tau} \exp(-\frac{s}{s_\xi}).$$

At the percolation threshold this expression can be approximated as

$$n_s(p) \propto s^{-\frac{5}{2}} exp(-\frac{s}{s_\xi}). \tag{1.13}$$

Therefore, the cluster size distribution follows a power-law with an exponential cutoff.

1.6 Summary

In this chapter we have reviewed some of the works that have been carried out in the last few years in the area of networks science. We have started by giving a general definition of the networks and showed that the recent work on the mathematics of the networks has been driven largely by observations of the properties of actual networks, ranging from communication networks to ecological webs. and attempts to model them. In general we have defined two types of networks. static networks and dynamic networks. In this thesis we will focus only on the static networks. Further discussion about these types of networks will be in chapters 3 and 4. We have then have reviewed some of the statistical properties

of networks that have received particular attention. We have particularly focused on three of these properties, namely, the degree distribution, average path length and the cluster size distribution. We then discussed how these properties can divide the networks into different types. In particular we discussed the similarities and contrasts between regular, random, small-world and scale-free networks. As we intend to link results obtained by the network science community to those obtained using percolation theory we also gave an introduction to the percolation theory, and discussed some quantities of interest in the field of percolation theory (percolation probability, average cluster size and cluster size distribution). These quantities can be compared to those of interest in random networks as well. Since the percolation has been solved exactly in one dimension, we reviewed this type of percolation, and studied all its properties. In fact, this type of percolation is corresponding to the some types of regular networks. We also studied the Bethe lattice, which has also been solved exactly, and its relation to infinite networks. There is an important difference between percolation theory and random networks. Percolation theory is defined on a regular d-dimensional lattice. In a random networks we can define the neighbourhood of a node as the set of nodes it is connected to. We can also define a non-metric distance along the edges to any other node in the network. There is no regular small-dimensional lattice a network can be embedded into. However, random network and percolation theory meet exactly in the finite dimensional limit ($d \to \infty$) of percolation. The infinite dimensional percolation is the same as the Bethe lattice, or, more precisely, the percolation threshold on which a giant cluster first forms on the Bethe lattice is exactly the same as the percolation threshold in random graphs (See also Chapter 3, Sec 3.2.2).

CHAPTER TWO

Percolation theory in the modelling of species distribution

2.1 Introduction

The main aim of this chapter[1] is to interpret the classical occurrence maps used in ecology through percolation concepts. Presence or absence maps of a given species, with N individuals distributed in a studied area A, are obtained by detecting empty and occupied cells in a regular square grid with cell area a (see *Figs. 2.1* and *Fig. 2.2*). a is the varying resolution or scale and A is the upper scale, where $0 < a < A$. If $p(a)$ is the probability for a cell of size a to be occupied, then $p(a)$ decreases from 1 to 0 when a decreases from A to 0. whatever the spatial distribution pattern. Species distribution patterns are generally classified as regular (human made plantation). random and aggregated. It is known that patterns in the distribution of individuals and species across space provide information critical to our ability to decipher the forces that structure and maintain ecological diversity [66]. We will describe some of common models that are used to model

[1]This chapter is based on our published paper [65]

the different patterns of distribution, and study the relationship between perimeter and abundance. and between the scale, as it is of particular significance. because they measure, to a great extent. the degree of fragmentation of landscapes [3]. The core of this chapter is the use of percolation theory to compare different patterns of species, particularly the random distribution and the aggregated distribution of species. The random distribution of species formed our case study. We used two different types of aggregated distribution; one has been used in [3] (hereafter we call it aggregated distribution type 1) and the other one is an arbitrary generated distribution (hereafter we call it aggregated distribution type 2).

2.2 Percolation Theory for the Distribution and Abundance of Species

Percolation theory has been applied in ecology to develop neutral models for landscape patterns. It provides a framework for examining landscapes as 2D grids. usually square grids of size $m \times m$. containing m^2 unique sites [67]. One of the key goals of ecology is to explain the distribution and abundance of species [68]. The number of individuals or the *abundance* of species in an area is a fundamental ecological parameter and a critical consideration when making management and conservation decisions [69]. The relationship between distribution and abundance of species is important because it is linked to fields such as biodiversity. The change in the biodiversity of a forest can be expressed as change in the distribution and abundance of all species occurring there. Macroecology can be

described as the study of emergent patterns of assemblages of species distributed over geographical spatial scales [70].

Understanding the relationship between distribution and abundance of species is a central concern of ecology and biogeography [3, 71, 72, 73]. To study this relationships, the distribution and abundance of species is typically measured by griding a study area [2] into a lattice and then enumerating the species in each cell of the lattice see *Figs. 2.1* to *2.4*, where species presence is represented by dots.

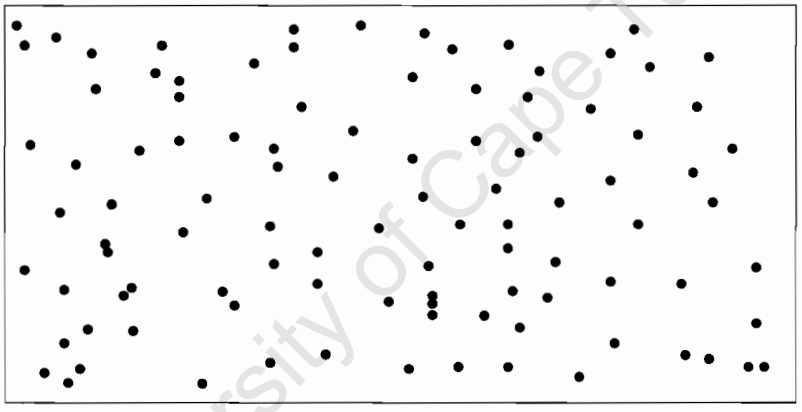


Figure 2.1: *A random distribution of dots in a study area. N = 100.*

[2]The area of occupancy of species is largely determined by three elements: the abundance of the species, its spatial distribution, and the cell size [3, 74]

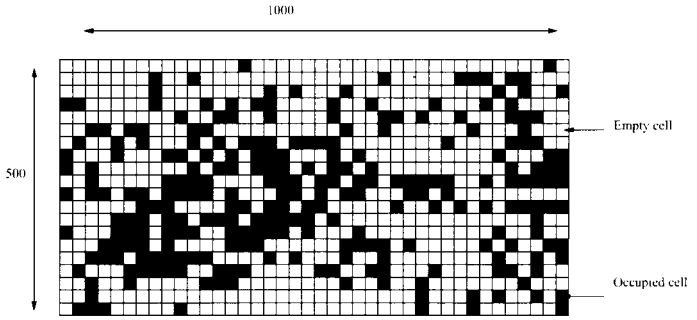


Figure 2.2: *The 2D square lattice with occupied (black) and empty (white) site (occurrence map), where* $a = 25 \times 25$, $A = 1000 \times 500$.

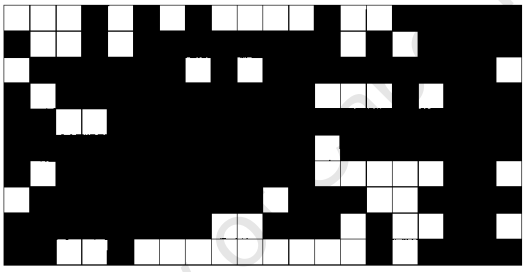


Figure 2.3: *The 2D square lattice with occupied (black) and empty (white) site (occurrence map), where* $a = 50 \times 50$, $A = 1000 \times 500$.

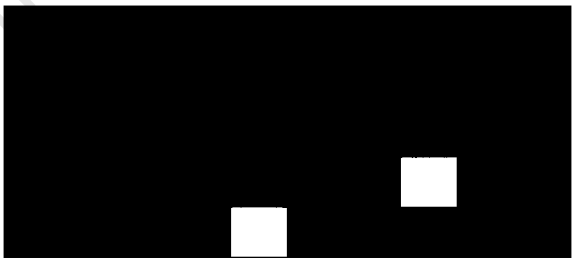


Figure 2.4: *The 2D square lattice with occupied (black) and empty (white) site (occurrence map) associated to Fig. 2.1 where* $a = 100 \times 100$, $A = 1000 \times 500$.

The distribution of species (dots) is recorded by occupancy (the number of occupied cells), while the abundance is recorded by the total number of individuals in all occupied cells, (in our study, the total number of dots in all occupied cells). From these figures one can build different percolation networks from the same data set depending on the grid scale [3].

2.3 The different patterns of distribution

In this section we give the definition of the three different patterns of *distribution* as follows:

1. Random distribution: This is a pattern of distribution in which species have an equal chance of occurring anywhere within an area, unaffected by the presence of other individuals (see *Fig. 2.5(a)*).

2. Regular (Uniform) distribution: In this distribution species are more or less uniformly spaced (see *Fig .2.5(b)*). They could even be exactly distributed on a regular grid as in some human-made forest.

3. Aggregated distribution: In this pattern species are more likely to be found in certain areas than others (see *Fig. 2.5(c)*).

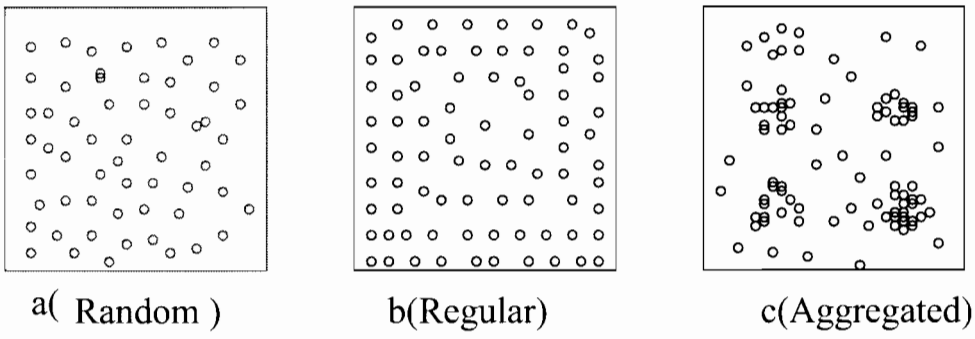


Figure 2.5: *a:Random distribution, b: Regular distribution. c: Aggregated distribution.*

2.4 Mathematical models for the patterns of distribution

In this section, we will consider random and aggregated distribution of species. There are a number of mathematical models being developed to quantify the relationship between distribution and abundance of species in a number of contexts [74. 75. 76]. In the subsections below, we review these models and point out the differences between the two types.

2.4.1 Random distribution for species

If individuals species are positioned randomly and independently of each other in space, this corresponds to the simplest natural occupancy-abundance model. As shown in *Figs. 2.1* and *Fig. 2.2*, each cell is either occupied (contains one or more than one dot) with probability p, or empty with probability $(1-p)$. Using a random distribution of dots, the probability of obtaining a particular dot in a sampling area cell size a is $\frac{a}{A}$. Therefore,

the number of dots n in each cell size a follows a binomial distribution [69]. The status of a cell to be occupied or not can be modelled by a Bernoulli trial. and the probability of a cell to be occupied is

$$p = 1 - \left(1 - \frac{a}{A}\right)^N, \tag{2.1}$$

where a is the cell size, A is the total area, N is the total number of dots in the study area.

2.4.2 Aggregated distribution of species

The random distribution of species does provide a general model of spatial pattern in ecology, and numerous studies have shown that this random distribution inadequately describes spatial patterns of aggregation of species across landscape [77, 78]. In nature, except at very low abundances. individuals of most species are typically aggregated, and the most frequently applied is the negative binomial distribution (NBD) [69]. The NBD is characteristic of many species distributions and can be used to describe regular, random or aggregated distributions [79]. To model aggregated species distribution, and to investigate the effect of aggregation on species-area relationship. a number of authors have examined the following model [80, 69, 74, 81]

$$p = 1 - \left(1 + \frac{aN}{Ak}\right)^{-k}, \tag{2.2}$$

where p is the probability of a cell to be occupied, and k is a clumping parameter describing the spatial dispersion of the species in the area. In general. k will be species and scale dependent [82]. $k \in (-\infty, -\mu) \cup (0, +\infty)$, where $\mu = aN/A$ is the mean density of the

species per cell [83, 3].

When k is positive, it corresponds to an aggregated pattern. The model derives from the negative binomial distribution (**NBD**), which has been used to describe species-area relationships and abundance aggregation patterns [69]. Small positive values of k represent stronger aggregation of species. Negative values of k describe a regular pattern of distribution. When the value of k is equal to the opposite of the number of species (*i.e.* $k = -N$), then *Eq. 2.2* is equivalent to *Eq. 2.1*. When $k \to \pm\infty$, the spatial distribution tends toward the Poisson distribution [83]. This is the case of our study. So the probability of a cell being occupied tends toward

$$p = 1 - e^{-\frac{Na}{A}} \tag{2.3}$$

where p is the probability of occurrence of species in a sample unit, and $\mu = \frac{aN}{A}$ is the mean local density of species (the mean density per sample). Therefore, when individuals of different species are distributed randomly on the landscape, there is an expected linear relationship between $-ln(1-p)$ and μ with a slope parameter of 1 [84]. For example, let's assume a total of N individuals of species randomly distributed among a given number of sites M of which m sites are occupied. Then,

$$p = m/M \quad and \quad \mu = N/M.$$

We used *Eq. 2.1* and *Eq. 2.3* for the cell being occupied, and, as we mentioned before, we will compare our calculation by using these two equations. *Eq. 2.2*, which describes the aggregated distribution, has been used in [3].

2.5 Perimeter definition in a 2D percolation model

In order to describe the spatial structure of species distribution, it may be important to consider the variation of the perimeter lengths and cluster sizes defined as follows:

1. The *perimeter L of the distribution* is defined as the sum of the length of the joints (boundaries) between occupied and empty cells at a given scale a, as illustrated by the 2D percolation models shown in *Figs. 2.2, 2.3* and *2.4*.

2. A *Cluster or (patch)* is a group of occupied cells which are connected side by side.

He And Hubbell's percolation theory (in [3]) is proposed as a solution to one of the major unresolved challenges in mathematical ecology, namely, the search for a reliable quantitative understanding of species abundances and distribution. They claim that they *have possibly unified several fundamental ecological parameters including abundance N, distribution of pattern k, scale a, extent A and edge length L into a single mathematical framework for the spatial architecture of fragmented metapopulations.* They establish a relationship between perimeter and abundance and between the number of patches and abundance. The main idea of the paper [3] that we studied intensively, was to look at the perimeter as a function of scale (cell size a) and abundance N. He and Hubbell (in [3]) assume that the occupancy of cells in a map are independent events. Under this assumption, we considered the probability for a joint between two cells to be an actual boundary between an occupied and an empty cell to be $2p(1-p)$, because a given join may separate occupied cells from empty ones in two ways. So the perimeter L has the

form:

$$L = 2J\sqrt{a}(1-p)p \tag{2.4}$$

Depending on the occurrence probability defined by *Eqs. 2.3*, *2.2* and *2.1*, the perimeters L will yield one of the following forms:

$$L = 2J\sqrt{a}\left(1 - e^{\frac{-Na}{A}}\right)e^{\frac{-Na}{A}} \tag{2.5}$$

$$L = 2J\sqrt{a}\left[1 - \left(1 + \frac{aN}{Ak}\right)^{-k}\right]\left(1 + \frac{aN}{Ak}\right)^{-k} \tag{2.6}$$

$$L = 2J\sqrt{a}\left[1 - \left(1 - \frac{a}{A}\right)^{N}\right]\left(1 - \frac{a}{A}\right)^{N}, \tag{2.7}$$

where J is the total number of neighbouring joins for a map [3]. For a regular lattice such as in *Fig. 2.2*, the value of J can be:

- $J = 2j_xj_y - j_x - j_y$, if we consider only the internal joins between cells, and
- $J = 2j_xj_y$, if we consider a toroidal boundary,

where j_x and j_y are the number of cells along x and y axis.

Eq. 2.5 to *Eq. 2.7* show that the perimeter L of distribution depends on the number of dots N, and the spatial scale (cell size a); but in *Eq. 2.6*, L depends also on the clumping parameter k. In each case, the perimeter L is a nonlinear function of the abundance. L first increases with N until it reaches a peak at a critical value N_c, after which L decreases with N (see *Fig. 2.6*). From the Percolation Theory [55] developed for random infinite networks, one knows that a percolation threshold p_c exists where a critical transition occurs. This threshold is of great potential interest in ecological

applications since it determines critical transitions and universal laws for any property linked to the connectivity of the filled space. For any given (N, A) and varying a, there exists a critical scale a_c, where $p(a) = p_c$, and a percolation-like transition occurs. In a similar way, for any given (A, a) and varying N, there exists a critical abundance N_c for percolation transition. This critical point N_c is called in [3], *percolation threshold*. In fact

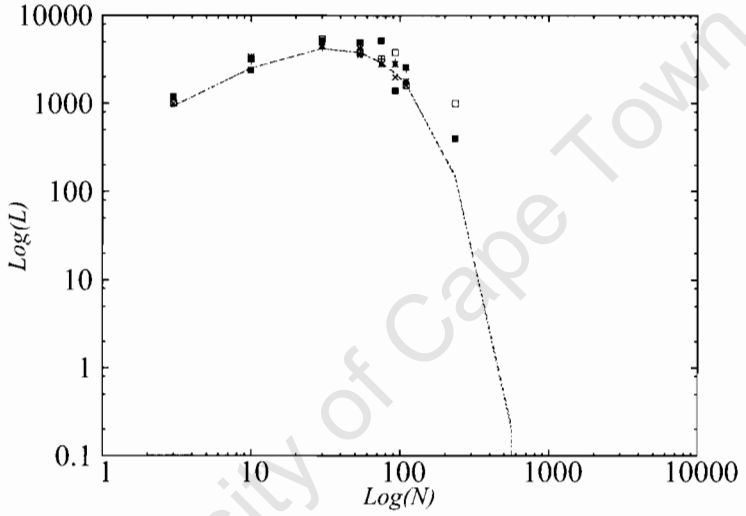


Figure 2.6: *Perimeter* (L)*-Abundance* (N) *relationship for random distribution, provided cell size* $= 100 \times 100$. *It shows that the value of* (L) *increases with* (N) *until reaching* N_c *after that it decreases. Each dot represent the actual perimeter value for the dots, and the curve is the predicted from Eq. 2.5.*

He and Hubble's obtained these critical values from *Eq. 2.5*. *2.6* and *2.7* respectively, by differentiating the function $L(N)$ and calculating the value N_c where L reaches its maximum. Depending on the selected probability distribution, one has:

$$N_c = -\frac{A}{a}\ln(1/2) \tag{2.8}$$

$$N_c = \left(2^{1/k} - 1\right) \frac{Ak}{a} \tag{2.9}$$

$$N_c = \frac{\ln(1/2)}{\ln(1 - a/A)} \tag{2.10}$$

2.6 Simulation and results

We used different values of N (number of dots), and converted the spatial distribution of each set of dots into an occurrence map for different scales. We counted the number of dots in each of the cells. We then calculated the value of the perimeter L, and compared the calculated value of the perimeter L with the predicted perimeter value of *Eq. 2.5* and *Eq. 2.7*. For each cell size a and number of dots N, we have five values for the calculated values of L (we generated our random dots five times to get these values) and usually two curves for the predicted values according to *Eq. 2.5* and *2.7*. We found that the values of perimeter L from these two equations are close to each other. Also for large values of N these two equations give approximately the same value of L.

In *Figs. 2.7* to *2.12*, we represent $\log(L)$ as function of $\log(a)$. The calculated values of the perimeter are represented by dots, and the two lines show the predicted values of L using *Eq. 2.5* and *Eq. 2.7*. We considered the two types of boundaries (See the Appendix for the code), a toroidal boundary (all the joins between the cells is $J = 2j_x j_y$), shown in *Figs. 2.7* to *2.9*. and the only internal joins between the cells ($J = 2j_x j_y - j_x - j_y$) , shown in *Fig. 2.10* to *2.12*.

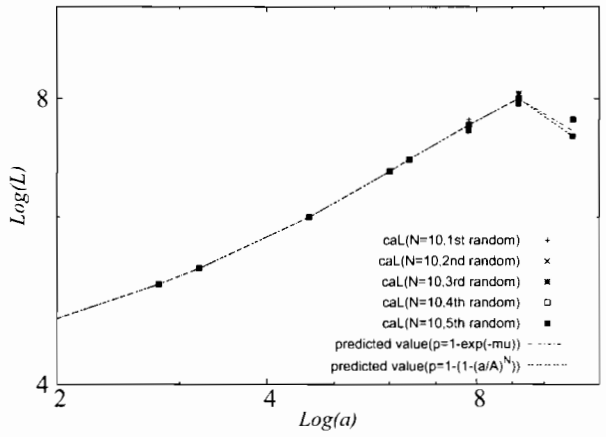


Figure 2.7: *Perimeter* (L)*-scale* (a) *relationship for random distribution,* $N = 10$ *across scales* 2×2, 4×4, 5×5, 10×10, 20×20, 25×25, 50×50, 100×100 *and* 250×250, *with toroidal boundary.*

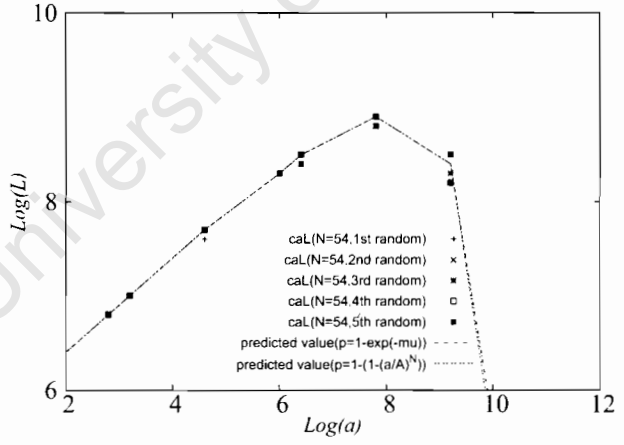


Figure 2.8: *Perimeter* (L)*-scale* (a) *relationship for random distribution,* $N = 54$ *across scale* 2×2, 4×4, 5×5, 10×10, 20×20, 25×25, 50×50, 100×100 *and* 250×250, *with toroidal boundary.*

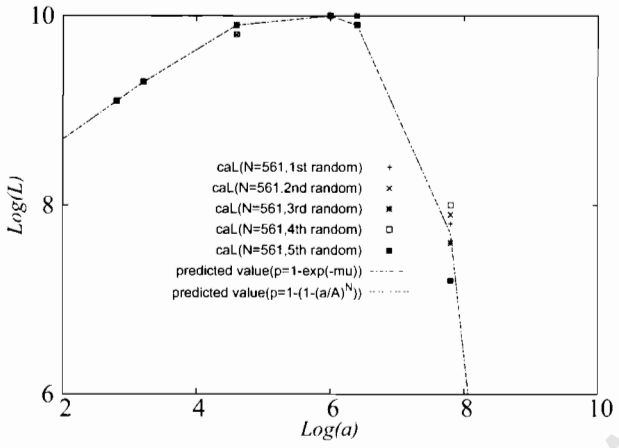


Figure 2.9: *Perimeter* (L)*-scale* (a) *relationship for random distribution,* $N = 561$ *across scales* 2×2, 4×4, 5×5, 10×10, 20×20, 25×25, 50×50, 100×100 *and* 250×250, *with toroidal boundary.*

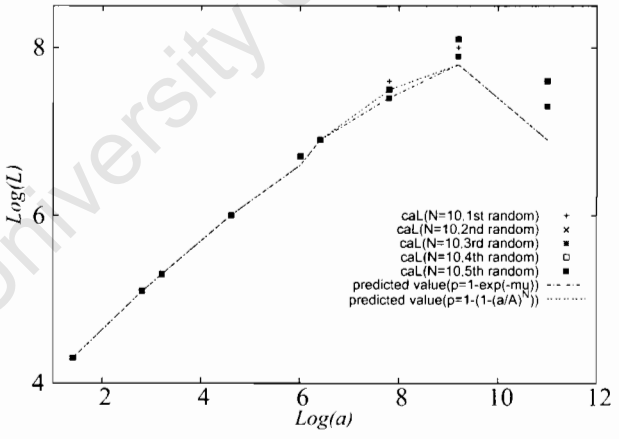


Figure 2.10: *Perimeter* (L)*-scale* (a) *relationship for random distribution,* $N = 10$ *across scales* 2×2, 4×4, 5×5, 10×10, 20×20, 25×25, 50×50, 100×100 *and* 250×250, *by considering only the internal joins between the cells.*

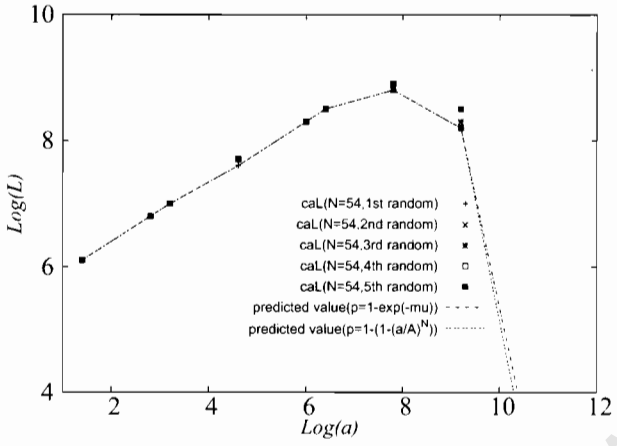


Figure 2.11: *Perimeter* (L)*-scale* (a) *relationship for random distribution,* $N = 54$ *across scales* 2×2*,* 4×4*,* 5×5*,* 10×10*,* 20×20*,* 25×25*,* 50×50*,* 100×100 *and* 250×250*, by considering only the internal joins between the cells.*

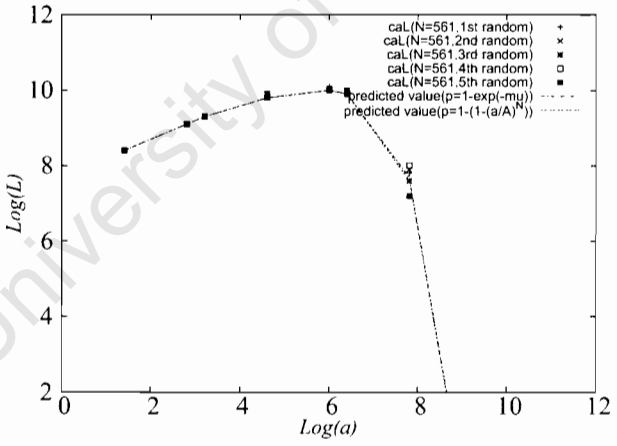


Figure 2.12: *Perimeter* (L)*-scale* (a) *relationship for random distribution,* $N = 561$ *across scales* 2×2*,* 4×4*,* 5×5*,* 10×10*,* 20×20*,* 25×25*,* 50×50*,* 100×100 *and* 250×250*, by considering only the internal joins between the cells.*

2.7 Comparison between random and aggregated distribution

In this section, we are going to compare the random distribution and the two different types of aggregated distributions.

2.7.1 Aggregated distribution of species type 1

2.7.1.1 Relationship between perimeter L and scale a

All the calculations in the previous sections were done for the random distribution of species. As we mentioned before, we will compare the random distribution of dots with the results given by [3] using aggregated pattern of distribution type 1 (Barro Colorado Island **(BCI)**). *Fig. 2.13* below shows the perimeter (L)-scale (a) relationship for **BCI** species across scales from 5×5 to $250 \times 250m$.

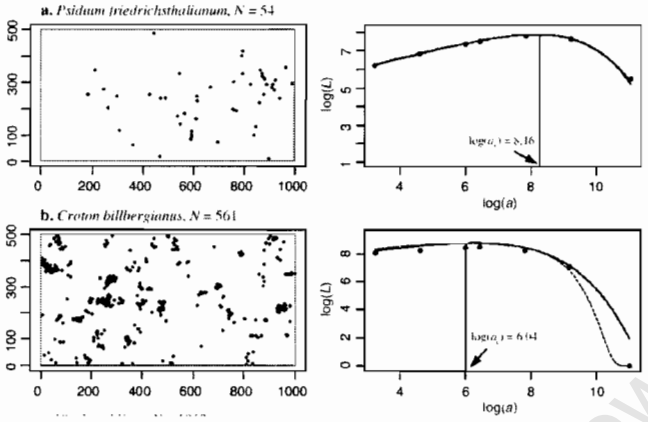


Figure 2.13: *Perimeter (L)-scale (a) relationship for BCI species. across scales from* 5×5 *to* $250 \times 250m$. *The left hand column shows the actual spatial distributions of the species. The observed perimeter data are shown on the right-hand column (dots), and the smooth curve is the prediction of Eq. 2.6. this Figure has been extracted from [3].*

The left hand column shows the actual spatial distribution of the species. The observed perimeter data is shown on the right hand column (dots) and the smooth curve is the prediction of *Eq. 2.6* using a scaling estimation function for (k) across scale. We can see that the results from *Fig. 2.7* to *2.12* (perimeter-scale relationship) fit very well to the random data generated by computer simulations, even for small values of N, and are very similar to those obtained in [3] using aggregated distribution type 1.

2.7.1.2 Relationship between perimeter L and number of species N

Figs. 2.14 to *2.16* show the relationships between perimeter L and the number of dots N with different cell size.

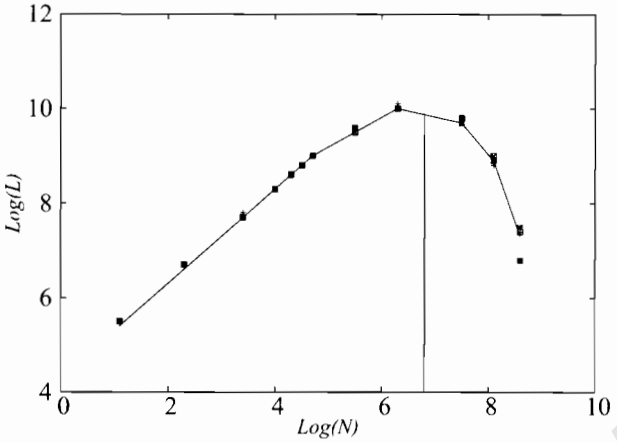


Figure 2.14: *Perimeter* (L)*-Abundance* (N) *relationship for random distribution of dots, provided* $a = 20 \times 20$. *The dots represent the actual distribution of the random species (dots), and the curve represent the prediction of Eq. 2.5*

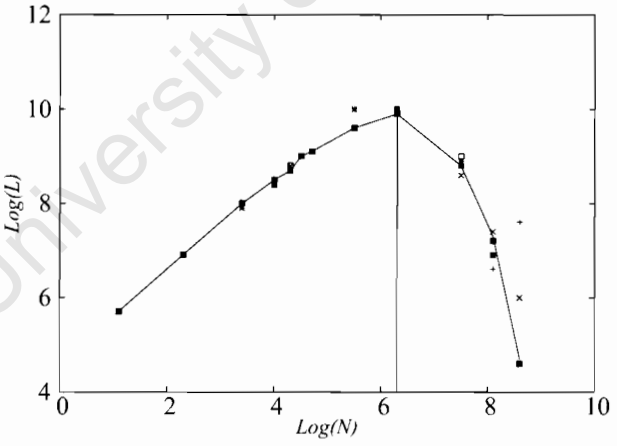


Figure 2.15: *Perimeter* (L)*-Abundance* (N) *relationship for random distribution of dots, provided* $a = 25 \times 25$. *The dots represent the actual distribution of the random species (dots), and the curve represent the prediction of Eq. 2.5*

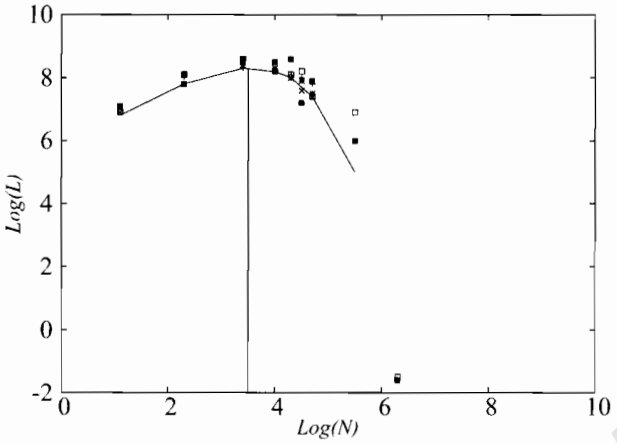


Figure 2.16: *Perimeter* (L)*-Abundance* (N) *relationship for random distribution of dots, provided* $a = 100 \times 100$. *The dots represent the actual distribution of the random species (dots), and the curve represent the prediction of Eq. 2.5*

In each figure we fixed the scale (cell size) and changed the total number of dots from $N = 3$ to$N = 5220$ for that scale. We used *Eq. 2.5* and *Eq. 2.7* to find predicted values of L. These figures show that the perimeter first increases with N until reaching a peak at a critical point N_c after which it decreases with N. Each dot in these figures represent the calculated value of L. The curves are predicted from *Eq. 2.5* and *Eq. 2.7*. The critical abundances N_c are calculated from *Eq. 2.8*.

As we did before, we compare the perimeter-abundance relationship between the random distribution of species (dots) and the aggregated pattern distribution type 1 done by [3]. *Fig. 2.17* shows the perimeter-abundance relationship for (**BCI**) species at four scales $5 \times 5, 10 \times 10, 25 \times 25$ and $50 \times 50m$.

Each dot represents the observed perimeter for the species: the smooth curves are

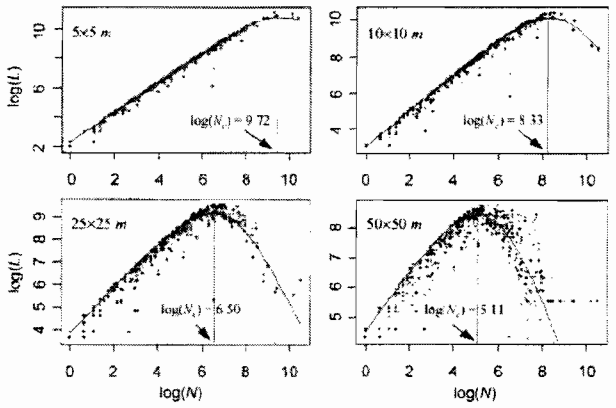


Figure 2.17: *Relation between the perimeter L and the abundance N for the* ***BCI*** *species at four scales $5\times5, 10\times10, 25\times25$ and 50×50. Each dots represent the observed perimeter for a species, the curves are the prediction of Eq. 2.6 using a universal $k = 2$ for all the species. This Figure is extracted from [3].*

predicted from *Eq. 2.7* using ($k = 2$) for all species. The critical abundances N_c has been calculated using *Eq. 2.9* at $k = 2$ and $A = 500000$.

We conclude that the curves are very similar for our randomly generated set of dots and for the type 1 of aggregation exhibited by the actual tree data considered in [3]. We also compared the value of the critical abundances obtained for the tree data and for the random distribution as a function of scale, using the same type of power law analysis that previously published in [3] (see *Fig. 2.18*) that is checking the linearity of $\log(N_c)$ versus $\log(a)$. We again conclude that there is no noticeable difference between random data and aggregated data of type 1 which could be observed using a percolation modelling approach.

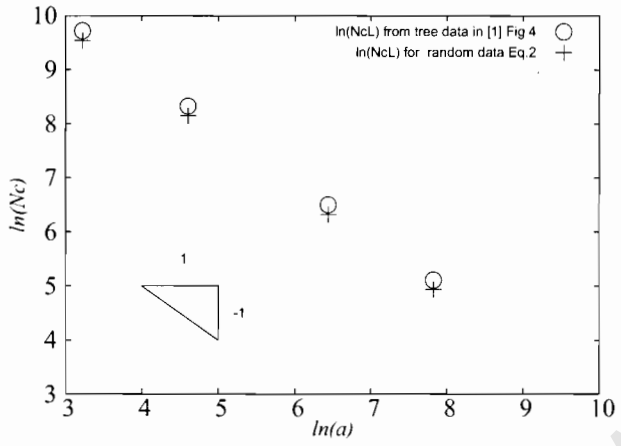


Figure 2.18: *Values of different critical abundances as a function of scale, for the tree data from Ref [3], Fig. 4 and Eq. 2 are in Ref [3], and for random distribution.*

2.7.2 Aggregated distribution of species type 2

Aggregated distributions could be generated in different ways. We generate an example of aggregated distribution of species (dots), as shown in *Fig. 2.19.* and we studied the perimeter-scale relationship.

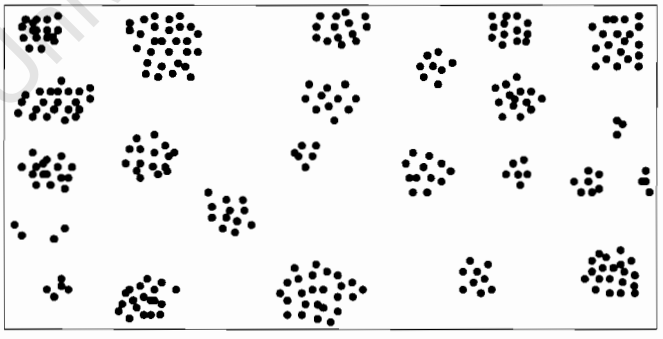


Figure 2.19: *Aggregated distribution of 400 dots in 2D.*

Figs. 2.20, 2.21 and *2.22* show the associated percolation models obtained at different scales, and the relationship between the perimeter L and the scale a.

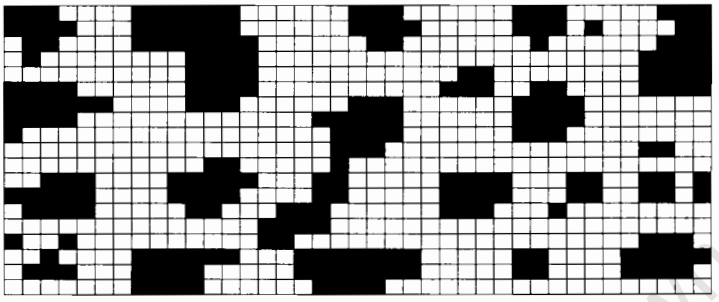


Figure 2.20: *The 2D square lattice with occupied (black) and empty (white) site (occurrence map) associated to Fig. 2.19 (is approximately associated, becuase it had done by hand), for the aggregated distribution, with* $a = 25 \times 25$, *with rectangle size* 1000×500.

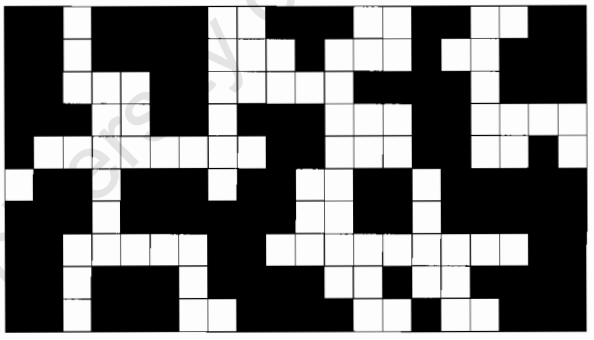


Figure 2.21: *The 2D square lattice with occupied (black) and empty (white) site (occurrence map) associated to Fig. 2.19, for the aggregated distribution, with* $a = 50 \times 50$, *with rectangle size* 1000×500.

Figure 2.22: *The 2D square lattice with occupied (black) and empty (white) site (occurrence map) associated to Fig. 2.19, for the aggregated distribution, with* $a = 100 \times 100$, *with rectangle size* 1000×500.

One can notice that the smaller the scale size, the bigger the perimeter value, then if we go on decreasing the scale size the value of perimeter remains approximately the same. In other words, the perimeter value does not change significantly, as shown in *Fig. 2.23*. This can be easily understood, because in this type of aggregation, if the density of the dots inside the cluster is very high, the occupied cells in the percolation model has to cover the same cluster area, whatever the scale a (unless the scale size is very small), and the perimeter remain constant and approximately equal to the perimeter of the cluster.

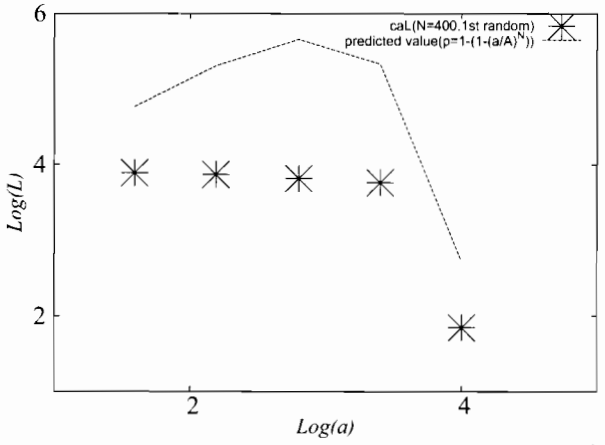


Figure 2.23: *Relationship between perimeter L and the scale size a. It shows that the smaller the size, the bigger the perimeter value, and by decreasing the scale size the perimeter value remains approximately the same.*

2.8 Conclusion

In this chapter we have presented the relationship between the perimeter and abundance, and between perimeter and scale using random patterns. We have also done a comparison between two different patterns of distributions, mainly, between the random distribution and aggregated distribution. We have done this by ensuring that all the equations describing the random distribution fit well with the random data (or what ecologists call the 'null hypothesis' or the "neutral model") generated by computer simulation, and comparing this case study with two different types of aggregated distribution, type 1 and 2. We have found that there is no difference between aggregated and random distribution which can be detected by using the definition of percolation threshold. when we compare the random distribution of species with the aggregated distribution type 1, but this is

not the case when we compare the random distribution with our aggregated distribution, type 2. We have shown that the relationship between perimeter L and scale a is not the same in the two cases. The underlying power law scaling relating a critical abundance value N_c and the scale a are identical in the case of random distribution and aggregated distribution type 1. We consider that this power law might be an intrinsic property of any spatial data due to mere scaling effects. We can conclude that percolation threshold may not be used to discriminate between different patterns; they are identical for random and aggregated distribution type 1. But it may depend on how the data is aggregated, as shown in our first exploration as regard the aggregated distribution type 2.

CHAPTER THREE

Comparison between 2D spatial lattice networks and random networks

3.1 Introduction

In this chapter we will consider two types of network, the random network denoted by **RN** [85, 86, 87] and a 2D spatial lattice network denoted by **SN** bulit from percolation models. Many of the average properties of the **RN** have been calculated exactly in the limit of large N, where N is the network size [88, 89]. Spatial networks are networks in which the way nodes are connected is defined according to spatial neighbouring properties (Euclidean distance between two nodes). Thus, spatial networks impose another constraint on the nodes within a network [90]. Here we will consider **SN** network defined as follows: In a $n \times n$ square grid, we define p_s the probability for a cell to be occupied and to represent a node of **SN**. Thus we have $\overline{N} = p_s n^2$. We will compare this type of network with a classical **RN** network, where p_r represents the probability between two arbitrary nodes within N to

be connected. The main purpose in this chapter is to examine the effect of varying p_r and p_s and to make a comparison between these two networks. To this end, we shall first revisit the properties of **RN** [91, 88, 4, 85], by generating **RN** with probability p_r of two nodes being connected. Then we shall look at a specific network (2D spatial lattice network **SN**). with probability p_s of cell being occupied, particularly emphasizing its spatial form. We will then develop algorithms to study the degree distribution, cluster size distribution and path length distribution of the two networks. Finally we will investigate the effect of varying the two probabilities for the two networks, when $p_r$andp_s are below or above critical percolation values p_{rc}andp_{sc}. We will also derive an original and general expression for the degree distribution of **SN** and for the average path length when $p_s = 1$.

3.2 Definition and construction of RN and SN

3.2.1 Random networks RN

Random network is the simplest network model, and it is a type of a static network, because the total number of nodes is fixed. The construction of a network is often called the evolution process: starting with a number of isolated nodes N, the graphs develops by the successive addition of random links [4]. We create **RN** as follows:

- Distribute N nodes $i = 0, 1, 2, ..., N$.
- Connect at random these nodes.

The above procedure provides random uncorrelated networks. From the mathematical point of view, the lack of correlations means that the probability $P(k_i/k_j)$, that an edge departing from a vertex j of degree k_j arrives at a vertex i of degree k_i, is independent of the initial vertex j [92]. *Fig. 3.1* shows an example of **RN**, where the probability of two nodes being connected is $p_r = 0.02$. As the probability of connection increases, we get a highly connected network (as we can see in *Fig. 3.2* $p_r = 0.09$), eventually obtaining a fully connected graph when $p_r = 1$.

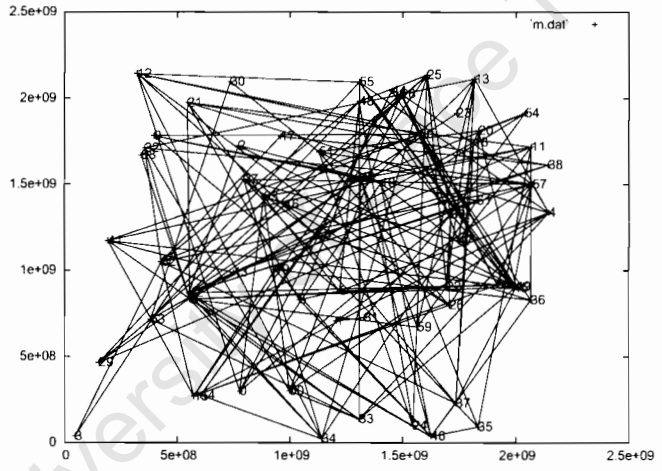


Figure 3.1: *A random network where the probability for two nodes being connected is* $p_r = 0.02$, $N = 100$.

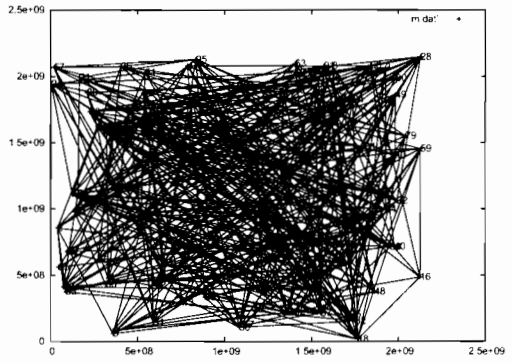


Figure 3.2: *A random network where the probability for a cell being occupied is* $p_r = 0.09$, $N = 100$.

Before revisiting the properties of this network we shall first define the use of the generating function.

3.2.1.1 Generating functions

In this section we describe the use of the generating function which was developed in the work by *Newman*, *Strogatz* and *Watts* [6]. A probability generating function $G_0(x)$ is an alternative representation of the probability distribution p_k, which is the degree distribution of vertices of degree k. It is defined as follows

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k, \tag{3.1}$$

and the distribution is assumed correctly normalized so that $G_0(1) = 1$. This function encapsulates all the information contained in p_k, since

$$p_k = \frac{1}{k!} \frac{d^k G_0}{dx^k}|_{x=0}. \tag{3.2}$$

The mean degree can be calculated directly by differentiation of the generating function, which identifies the average number of first neighbours of the network

$$z = \langle k \rangle = \sum_k k p_k = G_0'(1) \tag{3.3}$$

We can calculate any moment of the distribution by taking a derivative [6]. So in general,

$$\sum_k k^n p_k = \left[\left(x \frac{d}{dx} \right)^n G_0(x) \right]_{x=1}. \tag{3.4}$$

The usefulness of the generating function approach is that the distribution of the sum of n randomly chosen nodes of degree k is generated by $[G_0(x)]^n$.

3.2.2 2D spatial lattice networks, SN

The structure of 2D **SN** is as follows: First we grid the study area into n^2 small cells, after which we distribute nodes randomly in the centre of each cell. The probability that one cell has a node (occupied) or not (empty) is $p_s = \frac{N}{n^2}$, where N is the number of occupied cells, and n^2 is the total number of cells. We then link two cells (nodes) if they are sharing an edge. When we analyse properties of a network the way the dots (nodes) and lines (links) are drawn is, in most cases, irrelevant. All that matters is which pairs of nodes are connected and which are not [4]. However in our model the location of the nodes is important, because the type of connections between them depend on actual neighbourhood in a geometrical or geographical space.

We can describe our network structure consisting of $N = p_s n^2$ nodes as follows :

- Represent each cell containing node i with labeled vertex (node)ν_i, $i = 1, 2, ..., N$.

- Draw links between vertices (nodes) ν_i and ν_j if cells i and j are sharing an edge (see *Figs. 3.3* and *3.4*)

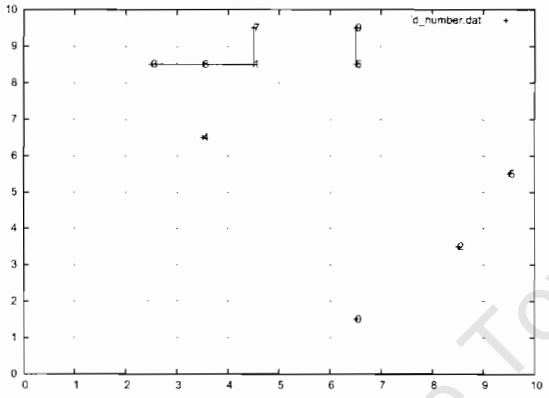


Figure 3.3: *A 2D spatial lattice network where the probability for a cell being occupied is* $p_s = 0.1$ $N = 10$.

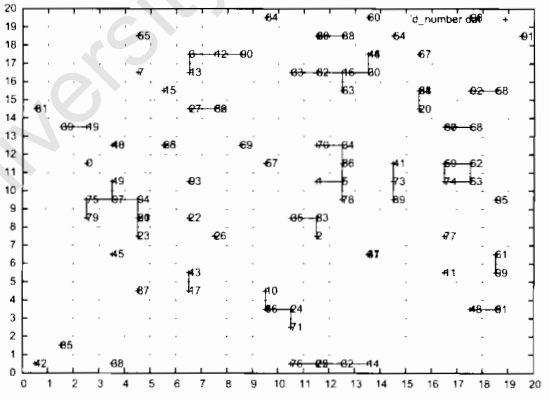


Figure 3.4: *A 2D spatial lattice network where the probability for a cell being occupied is* $p_s = 0.4$, $N = 100$.

3.3 Networks properties and computer algorithms

In this section we are going to discuss degree distribution, cluster size distribution and average path length algorithms.

3.3.1 Degree distribution Algorithm

The simple local characteristic of a vertex is its degree [93]. The number of links a node has is called the node degree. As we mentioned before, not all nodes in the network have the same number of links and so there is spread in the number of links associated with a given node. This spread of node degree is characterised by a distribution function $P(k)$. The degree distribution is obtained by counting the number of nodes that have degree $k = 1, 2, 3, ...$ edges and dividing it by the total number of nodes N [94] (see *Fig. 3.5*).

3.3.1.1 Algorithm

1. Label all the nodes from $i = 0$ to N.
2. Start from node i and look for its neighbours (a list of neighbours).
3. Count the number of neighbours to this node.
4. Set $i = i + 1$ and go back to step 2.
5. Calculate the degree distribution for the whole network for a given p, and the mean degree value.

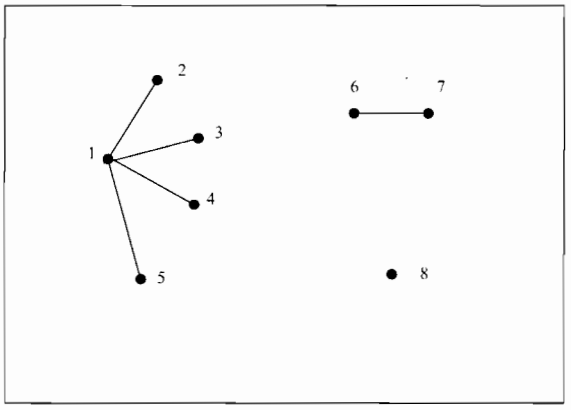


Figure 3.5: *An illustration of the definition of the degree, cluster and path length. This network has $N = 8$ nodes. 1 node has degree 0, 6 nodes have degree 1, 1 node has degree 4. There is 1 cluster of size 1, 1 cluster of size 2 and 1 cluster of size 5. The path length (2,5) = 2 and path length (5,8) = ∞ (no path between node 5 and node 8).*

3.3.1.2 Example

Below we give an example of the use of the degree distribution algorithms for the **RN**, where the probability for two nodes to be connected is $p_r = 0.02$, and another example for **SN** where the probability for a cell being occupied is $p_s = 0.1$.

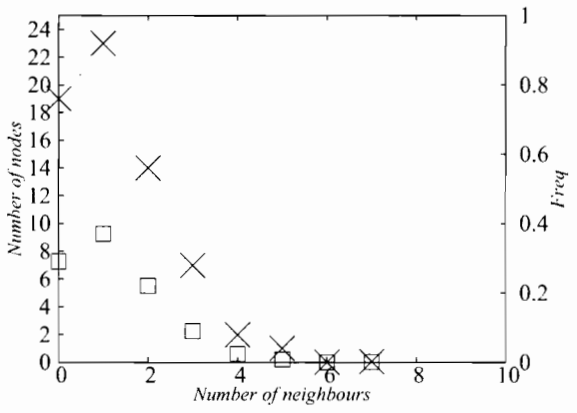


Figure 3.6: *Degree distribution for* ***RN*** *with the probability of two nodes being connected* $N = 64$, *when* $p_r = 0.02$. *The mean degree* ~ 1.22

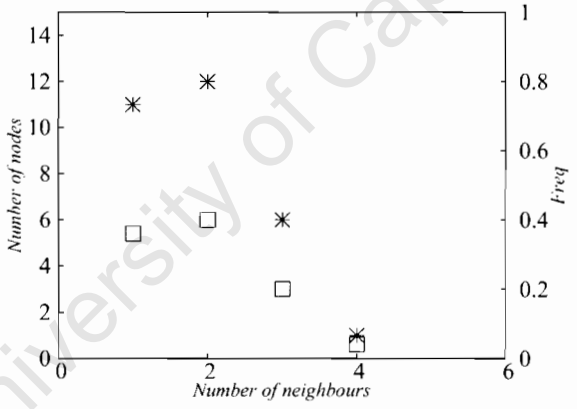


Figure 3.7: *Degree distribution for* ***SN*** *with* N *varies between 31 and 45, and total cell number* $n^2 = 64$, *when* $p_s = 0.1$, *squares represent the node number frequencies. The mean degree* ~ 1.92

3.3.2 Cluster size distribution Algorithm

A cluster is a subset of nodes in the network such that, for any two nodes $v_i$andv_j in the cluster, there exists a path connecting v_i to v_j (that is a group of occupied sites connected

by nearest neighbours distances for the **SN** network). See *Fig. 3.5.*

3.3.2.1 Algorithms

1. Label all the nodes from $i = 0$ to N.
2. Start from node i and look for its neighbours (a list of neighbours).
3. Look for the neighbours of the neighbour of the original node.
4. Stop if the last neighbour of the original node i has no more neighbours.
5. Count this list as one cluster of size s (s is the list length).
6. Set $i = i + 1$ and go back to step 2.

3.3.2.2 Example

Here we give two examples; one for the use of the cluster size distribution algorithms for the **RN** where the probability for two nodes to be connected is $p_r = 0.02$, and another example for **SN** where the probability for a cell being occupied is $p_s = 0.4$.

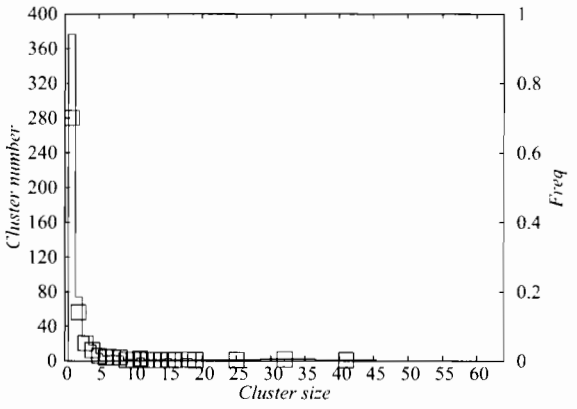


Figure 3.8: *Cluster size distribution for* ***RN****,* $N = 64$*, with probability that of two nodes being connected is* $p_r = 0.02$*, with mean cluster size* ~ 2.16*.*

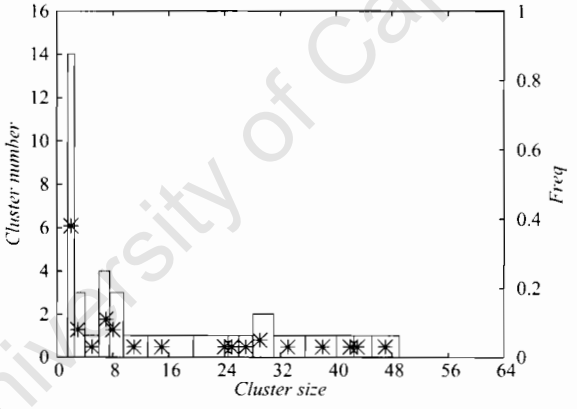


Figure 3.9: *Cluster size distribution for* ***SN*** *with* N *varies between 31 and 45 nodes, and total cell number* $n^2 = 64$*, counted over 10 iterations. With mean cluster size* ~ 13*.*

3.3.3 Path length distribution algorithm

Two nodes of a graphs are connected if a sequence of adjacent nodes. a path, links them. Path length between two nodes is defined as the number of edges along the shortest path connecting them [94]. We used the Dijkstra algorithm to find the average shortest path

length between two nodes in a network. It functions by constructing a shortest path tree from the initial node to every other node in the network (see *Fig. 3.5*).

3.3.3.1 Algorithm

- Label all the nodes from $i = 0$ to N.
- Start from node i and looking for its neighbours (a list of neighbours).
- For each node in the network look for the neighbours of the first neighbour.
- If it is connected to other node append this node to the list.

  If not, go to the next neighbour in the list, and so on.
- Take the minimum distance between two nodes in the network.
- Compute the distribution over all pairs of nodes in the network.
- Compute the average shortest path length.

3.3.3.2 Example

Here we give example of the use of the average path length algorithms for **SN** where the probability for a cell being occupied is $p_s = 0.85$. The average path length below the percolation threshold is unbounded.

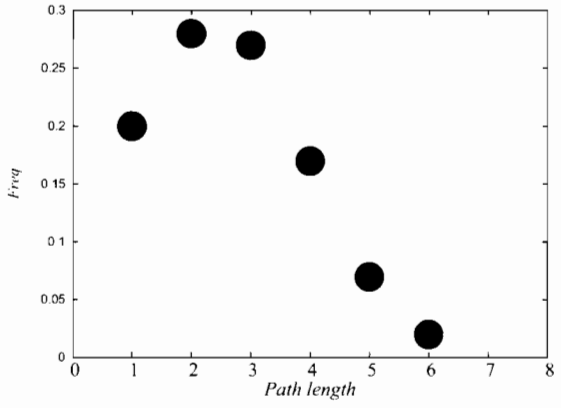


Figure 3.10: *Path length distribution for* ***SN****,* $n^2 = 16$ *when* $p_s = 1$ *using the path length distribution algorithm.*

3.4 Properties of Random networks

3.4.1 Degree distribution for RN

The degree k_i of node i in random network with probability of two nodes being connected p_r follows a binomial distribution. For a large N, the degree distribution of **RN** limits to Poisson distribution [1].

$$P(k) = \frac{e^{\langle k \rangle} z^k}{k!} \tag{3.5}$$

The generating function $G_0(x)$ in **RN** is given by [6]

$$G_0(x) = (1 - p + px)^N = e^{z(x-1)}, \tag{3.6}$$

where the last equality applies for a large N. The average degree from this equation is $G_0'(x) = \langle k \rangle = z = p_r(N - 1)$.

We used the above algorithm to calculate the degree distribution for **RN**, *Fig. 3.11* and

3.12 show the degree distribution for **RN** with probability of two nodes being connected is $p_r = 0.02$ and $p_r = 0.09$ using the algorithm defined above, with mean degree ~ 1.22 and ~ 5.49, respectively, as we can notice that, the high probability of two nodes being connected in **RN**, the high mean degree.

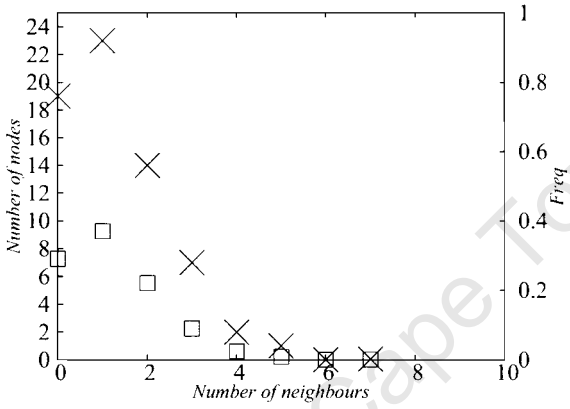


Figure 3.11: *Degree distribution for a random network with ($N = 100$, Fig. 3.1) probability of two nodes being connected as $p_r = 0.02$. The mean degree ~ 1.22*

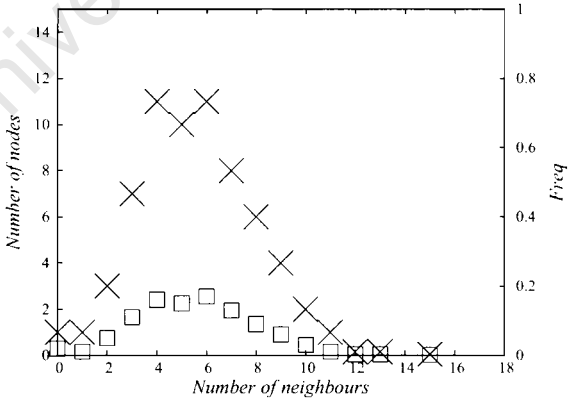


Figure 3.12: *Degree distribution for a random network ($N = 100$, Fig. 3.2) with probability of two nodes being connected as $p_r = 0.09$. The mean degree ~ 5.49*

3.4.2 Cluster size distribution for RN

The characteristics of the cluster size distribution below and close to the critical point have been studied widely [4, 88]. In this section, we discuss some of the main results of the cluster size distribution in **RN** below and close to the critical point (percolation phase transition p_{rc}).

Giant component and percolation threshold

The expected structure of **RN** varies as the value of p_r changes. The percolation threshold in **RN** given by $p_{rc} = \frac{1}{(N-1)}$, and for large N, is equal to $\frac{1}{N}$, this percolation threshold is the same as in the Bethe lattice (see Chapter 1). When $p_r < p_{rc}$ the network falls apart into small pieces, or in another word, the random network consists of small components. The probability that there is a path joining two nodes decays exponentially [4, 1], on the other hand when $p_r > p_{rc}$ in which most of nodes are joining together in a single giant component emerges [?], and the reminder of the nodes occupying smaller components also with again an exponential size distribution [1]. The expected size of the giant component is given by (the derivation of this equation is in [4])

$$S = 1 - e^{-zS} \tag{3.7}$$

Cluster size distribution

By using the cluster finding algorithm discussed above, we studied the cluster size distribution for **RN**. The algorithm we used is implicitly incorporated into the generating

function method [4]. For any **RN** with arbitrary degree distribution, let $H_1(x)$ be the generating function of the sizes of components which are reached by choosing a random edge. Following it to one of its ends, and it satisfies the iterative equation [95, 4]

$$H_1(x) = xG_1(H_1(x)). \tag{3.8}$$

If we start at a randomly chosen node, then we have one such component (cluster) at the end of each edge leaving that node, the generating function for the size of the whole component (cluster) will be

$$H_0(x) = xG_0(H_1(x)). \tag{3.9}$$

As *Newman et al* mentioned in [95], it is not usually possible to find a close expression for the complete cluster size distribution on a network, but they have been able to find a close expression for the average cluster size. This is given by

$$\langle s \rangle = H_0'(x) = 1 + \frac{G_0'(1)}{1 - G_1'(1)} \tag{3.10}$$

which diverges when $G_1'(1) = 1$. It is a mark of the phase transition at which a giant component (cluster) first appears. Substituting the definition of $G_0(x)$, one can write the condition for the first appearance of the giant cluster as

$$\sum_k k(k-2)p_k = 0 \tag{3.11}$$

from the definition of the giant cluster, *i.e.* a cluster containing a fraction S of nodes, thus, $H_0(1)$ is no longer unity, but it takes the value $1 - S$. So the size of the giant cluster is

$$S = 1 - G_0(u), \tag{3.12}$$

where u is the smallest non-negative real solution of $G_1(u)$. Close to the phase transition, the tail of the cluster size distribution n_s behaves as

$$n_s \sim s^{-\tau} e^{-s/s^*} \tag{3.13}$$

where the constant τ and s^* can be calculated from the properties of $H_0(x)$.

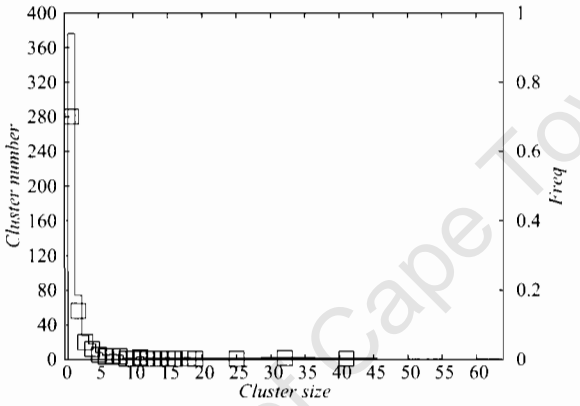


Figure 3.13: *Cluster size distribution for random network with the probability of two nodes being connected is* $p_r = 0.02$. *With mean cluster size* ~ 2.16.

Figs. 3.13 and *3.14* show the cluster size distribution for **RN** with the probability of nodes being connected as $p_r = 0.01$ and $p_r = 0.02$. The mean cluster size ~ 1.73 and ~ 2.16, respectively, using the above mentioned algorithm. We can notice that, for small values of p_r the network consists of isolated clusters with small sizes. and when $p_r < p_{rc}$, the largest cluster in **RN** is a tree, while for $p_r > p_{rc}$, the largest cluster contains cycles.

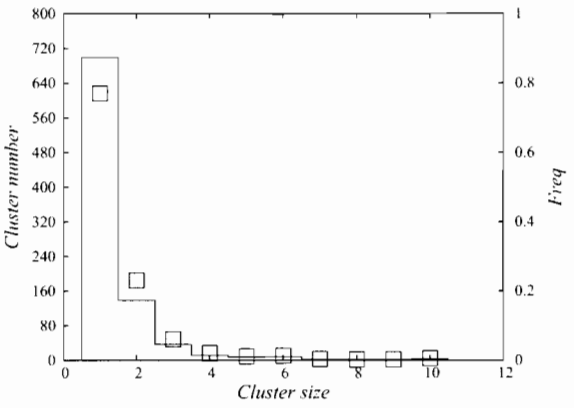


Figure 3.14: *Cluster size distribution for random network with the probability that a cell is being occupied is* $p_r = 0.01$. *With mean cluster size* ~ 1.73. $p_{rc} \approx 0.01$, above this value a giant component will appear.

3.4.3 Path length distribution for RN

It is well known that **RN** have a very small average path length, which scales as $l \sim \ln N$, where N is the network size [6, 21, 96]. As the average path length in the network gets smaller, the network becomes more connected. We can use the generating function formalism to calculate the average path length in **RN** with arbitrary degree distribution as follows: The probability distribution for the first and second neighbours can be generated by the functions $G_0(x)$ and $G_0(G_1(x))$. The average number of m^{th} neighbours will therefore be

$$z_m = [G_1'(1)]^{m-1} G_0'(1) = \left[\frac{z_2}{z_1}\right]^{m-1} z_1. \tag{3.14}$$

where, z_1 and z_2 is the number of first and second neighbours [6]. From the last equation, *Newman et al* estimated the length l of the shortest path between two random nodes on

the network. By assuming that all the nodes in the graphs can be reached within l steps (the total number of neighbours of a node out to that distance is equal to the number of nodes on the graphs), then we have

$$1 + \sum_{l}^{m=1} z_m = N. \tag{3.15}$$

However in most networks, this conditions holds $N >> z_1$ and $z_2 >> z_1$ we obtain [6]

$$l = \frac{\log(N/z_1)}{\log(z_2/z_1)} + 1. \tag{3.16}$$

A better approximation to l is to replace N in *Eq. 3.15* by SN, where S is the fraction of the network occupied by the giant cluster. The last result in *Eq. 3.16* reflects several generals properties of the average path length [4]:

- l scales logarithmically with N for all **RN**, irrespective of the degree distribution.
- Since the average path length l is a global property, it can be calculated from local properties as the average number of first and second nearest neighbours.
- It has been shown that only the average number of first and second nearest neighbours are important in the calculation of average path length l. Thus for two networks with completely different distributions but the same values of z_1 and z_2 we will have the same average path length.

3.5 Statistical properties of 2D spatial lattice networks

3.5.1 Degree distribution for SN

As we mentioned before this kind of network has been studied in the context of percolation theory, where results have been obtained as regards the percolation threshold p_{sc}, appearance of a giant cluster and fractal distribution of the cluster size distribution at p_{sc}. It has not been studied so far in the context of networks science and we will study also its mathematical properties such as, degree distribution and the path length distribution. In this section, we will be able to study the degree distribution for this kind of network analytically in the simple regular case when $p_s = 1$, then when $p_s \neq 1$, we will use our computer algorithm, and finally we will be able to calculate analytically the degree distribution, generalizing the regular case.

3.5.1.1 Degree distribution for fully connected SN, $p_s = 1$

In the case of fully connected **SN**, we are able to derive a general formula for the degree distribution. We start by giving a simple example. Let us consider a lattice of size 4×4, with a total number of 16 nodes, see *Fig. 3.15*, from this figure one can notice that the minimum node degree is 2. The maximum node degree is 4, as it will be the case for **SN**.

Thus, in the case of fully connected **SN**, we will have nodes with degree 2, 3 and 4 only (k = 2,3 and 4) where k is the node degree. The following equation describes the

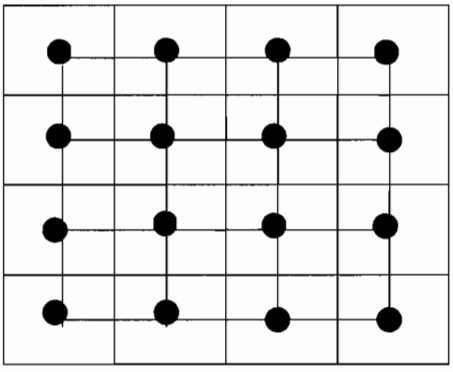


Figure 3.15: *Illustration of* 4×4 *fully connected 2D spatial lattice random network.*

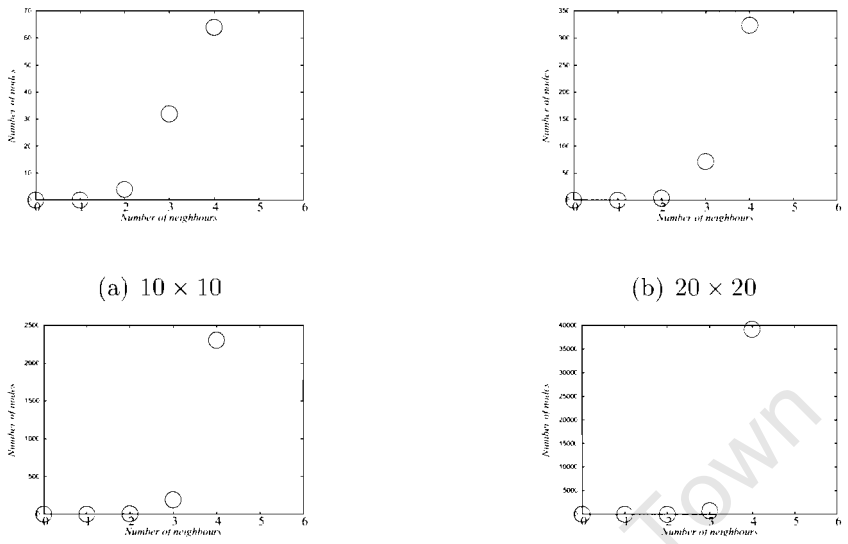
Table 3.1: *Number of nodes and the node degree for fully connected* ***SN***

Nodes degree	Number of nodes
2	4
3	8
4	4

degree distribution for this case

$$P(k) = \begin{cases} 0 & \text{for} \quad k = 0 \\ 0 & \text{for} \quad k = 1 \\ \frac{4}{n^2} & \text{for} \quad k = 2 \\ \frac{4(n-2)}{n^2} & \text{for} \quad k = 3 \\ \frac{(n-2)^2}{n^2} & \text{for} \quad k = 4 \end{cases} \tag{3.17}$$

The mean degree in this case is $\lambda = \frac{4(n-1)}{n}$. *Fig. 3.16* below shows the degree distribution for this case. with different lattice sizes.



(a) 10×10 (b) 20×20

(c) 50×50 (d) 200×200

Figure 3.16: *Degree distribution for fully connected 2D spatial lattice network with different lattice size.*

3.5.1.2 Degree distribution for SN when $p_s \neq 1$

As we know from percolation theory [55], by increasing the probability of a cell to be occupied (thus increasing the number of nodes in the network), all the nodes in the network forming a giant cluster, and by decreasing this probability the network will consist of isolated nodes, so the degree distribution in the case when $p_s \neq 1$ will be different from the fully connected network ($p_s = 1$). *Fig. 3.17* and *3.18* below show the degree distribution for **SN** $p_s = 0.1$, and $p_s = 0.9$ and mean degree ~ 1.92 and ~ 1.19.

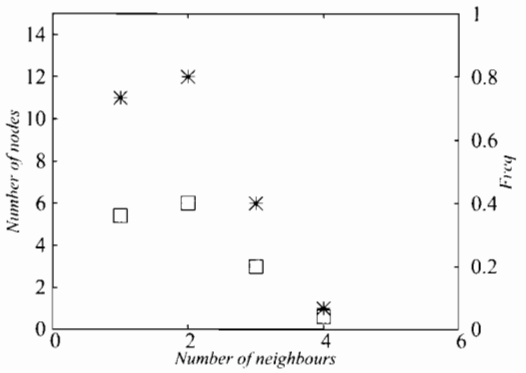


Figure 3.17: *Degree distribution for N between 31 and 45 random connected nodes with total cell number 64. $p_s = 0.1$. Squares represent the node number frequencies. The mean degree* ~ 1.92

Figure 3.18: *Degree distribution for N between 31 and 45 random connected nodes with total cell number 64. $p_s = 0.9$. Squares represent the node number frequencies. The mean degree* ~ 1.19

To do the mathematical calculation we considered two different cases to calculate the degree distribution for **SN** when $p_s \neq 1$. The first case is. neglecting the boundaries and assuming a very large network. We have calculated the degree distribution in this case

by calculating all the nodes with degree $0, 1, 2, 3$ and 4 as follows: To have nodes with degree zero, all its neighbours must be empty; to have nodes with degree one, three of its neighbours must be empty and only one is occupied; to have nodes with degree two, two of its neighbours must be empty, and two are occupied, to have nodes with degree three, one of its neighbours must be empty and three are occupied, and to have nodes with degree four. all the neighbours must be occupied. *Eq. 3.18* describes all these cases of the degree distribution when we neglect the boundaries of the network.

$$P(k) = \begin{cases} (1-p_s)^4 & \text{for} \quad k = 0 \\ 4p_s(1-p_s)^3 & \text{for} \quad k = 1 \\ 6p_s^2(1-p_s)^2 & \text{for} \quad k = 2 \\ 4p_s^3(1-p_s) & \text{for} \quad k = 3 \\ p_s^4 & \text{for} \quad k = 4 \end{cases} \tag{3.18}$$

In this case, we can easily calculate the mean degree, which is $\sum_{k=0}^{4} p(k)k = 4p_s$. For a very large network (where the size of the network $\to \infty$) and when $p_s = 1$ the above equation describe the behaviour of the degree distribution of the fully connected case which is. all the nodes have four degrees. This is given by $P(4) = p_s^4 = 1$.

The second case we studied for the degree distribution for **SN** when $p_s \neq 1$ is considering the boundaries of the network and assuming a very large network. The calculation of this is as in the previous case. In this case, all nodes could be either on a corner, on a boundary or inside the network. So to have nodes with degree zero, these nodes are either a corner and have zero neighbours, or on a boundary but not a corner with zero neighbours. or inside the network and also have zero neighbours. *Eq. 3.19* shows the

probability of having nodes with degree zero, and the same for the other degrees. The degree distribution in this case has the probability mass function (see *Eq. 3.19*).

$$P(k) = \begin{cases} \frac{(n-2)^2}{n^2}(1-p_s)^4 + \frac{4(n-2)}{n^2}(1-p_s)^3 + \frac{4}{n^2}(1-p_s)^2 & \text{for} \quad k=0 \\ \frac{(n-2)^2}{n^2}4p_s(1-p_s)^3 + \frac{4(n-2)}{n^2}3p_s(1-p_s)^2 + \frac{4}{n^2}2p_s(1-p_s) & \text{for} \quad k=1 \\ \frac{(n-2)^2}{n^2}6p_s^2(1-p_s)^2 + \frac{4(n-2)}{n^2}3p_s^2(1-p_s) + \frac{4}{n^2}p_s^2 & \text{for} \quad k=2 \\ \frac{(n-2)^2}{n^2}4p_s^3(1-p_s) + \frac{4(n-2)}{n^2}p_s^3 & \text{for} \quad k=3 \\ \frac{(n-2)^2}{n^2}p_s^4 & \text{for} \quad k=4 \end{cases} \tag{3.19}$$

We can easily prove that $P(k)$ in *Eq. 3.19* is a probability mass function, that is $\sum_0^4 P(k) = 1$. When $p_s = 1$ *Eq. 3.19* describe the case of fully connected network which has been given in *Eq. 3.17*. When the network is very large that is, $N(n^2) \to \infty$, *Eq. 3.19* describe the case of the degree distribution when we neglect the boundaries which has been given in *Eq. 3.18*.

3.5.2 Cluster size distribution for SN

Much has been investigated regarding the cluster size distribution as far as percolating clusters are concerned [97]. Cluster size ($n_s(p)$) is the total number of s-clusters (clusters containing s nodes). One can predict the behaviour of the cluster size of this type of network. Unfortunately, it is not possible to obtain an exact form for the cluster size $n_s(p)$ in $d > 1$ because there are very many different ways in which clusters can arrange themselves [55]. The probability that an arbitrary node belongs to an s-cluster is sn_s (every node must belong to one cluster since an isolated node surrounded by empty neighbours is also a cluster of size one). The probability that the node belongs to any

cluster is $\sum_{s=1}^{\infty} sn_s$, and the probability that a cluster to which a node belongs contains s-nodes is [55]

$$w_s = \frac{sn_s(p)}{p} = \frac{sn_s(p)}{\sum_{s=1}^{\infty} sn_s(p)} \tag{3.20}$$

When $p_s << p_{sc}$, **SN** consists of small and isolated clusters, thus n_s has contribution only from small sizes. As we approach p_{sc}, the size of the clusters in the network increase, until we find all the nodes in the network forming one giant cluster. Now there is contribution from all cluster sizes in the network. Thus, the mean cluster size $S(p)$ is given by

$$S(p) = \sum_{s=1}^{\infty} sw_s = \sum_{s=1}^{\infty} \frac{s^2 n_s(p)}{\sum_{s=1}^{\infty} sn_s(p)} \tag{3.21}$$

Fig. 3.19 shows the cluster size distribution for **SN** with mean cluster size ~ 12.

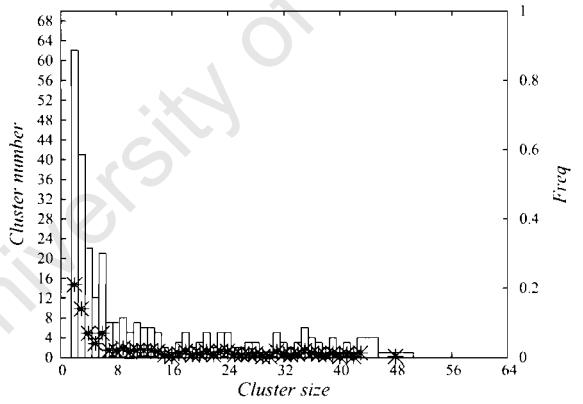


Figure 3.19: *Cluster size distribution for (31 $\sim$ 45) nodes with total cell number = 64, counted over 100 iterations. With mean cluster size* ~ 12.

We can define the correlation function $g(r)$ as the probability that if one node is in a cluster, then another node at distance r away is in the same cluster. This, typically, has

an exponential decay given by a correlation length ζ [55]

$$g(r) \rightarrow e^{-\frac{r}{\zeta}} \tag{3.22}$$

Eq. 3.22 indicates that the radius of the finite clusters has an exponential decay tail when $p_r < p_{rc}$, and the correlation length represents the mean radius of finite cluster. The exponential decay of a cluster implies that the probability that a cluster has size s also decays exponentially for large s [4], and the largest clusters have a fractal structure. When $p_r > p_{rc}$, there is exactly one infinite cluster, and the decay of the cluster size follows exponential, and in this case depends on the dimension of the lattice. However, as $d \rightarrow \infty$, the cluster size distribution decays exponentially, as in the case when $p_r < p_{rc}$ [4].

3.5.3 Path length distribution for SN

As we did in the degree distribution in **SN**, we will start first by deriving a general formula for the fully connected case $p_s = 1$ for the average path length. From *Fig. 3.15* one can notice that the minimum length in the network is one, and the maximum length between any two nodes on the network is $2(n-1)$. By considering the same example above, one can find all the possible paths between any two nodes on the network with the same length (The total number of possibilities of path length is $n^2(n^2-1)/2$).

The general formula of the average path length for **SN** is given as follows:

$$l(s) = \frac{4}{n^2(n^2-1)} \sum_{k=1}^{s} (n-k)(n-s+k), \qquad s-n < k < n. \tag{3.23}$$

Table 3.2: *Path length, # paths and the frequencies for fully connected* ***SN***

Path length	Number of paths	Frequencies
1	24	0.2
2	34	0.28
3	32	0.27
4	20	0.17
5	8	0.07
6	2	0.02

(*Fig. 3.20*) shows the average path length for **SN**, $p_s = 1$ using this formula and the path length algorithm.)

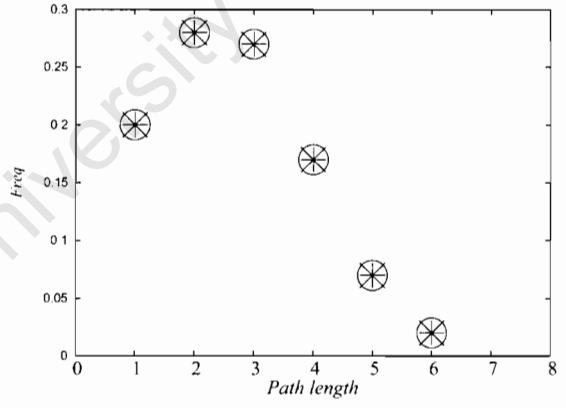
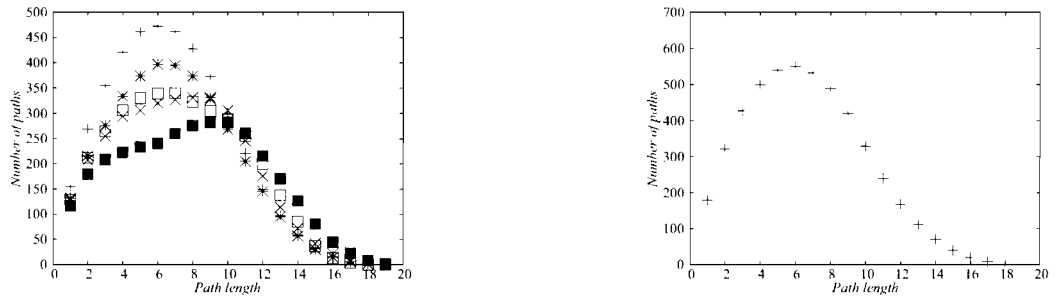


Figure 3.20: *Path length distribution for* ***SN***, $n^2 = 16$ *when* $p_s = 1$ *using Eq. 3.23 and the path length algorithm.*

For fully connected **SN**, the average path length is always bounded, but as soon as we have holes on the network, the average path length becomes unbounded. This is because there will be a number of nodes on the network which are not connected, hence there is no length between them. We consider this length as infinity. To avoid the unbounded values of the path length, *Neman* in [6] defines the path length in this case to be the mean distance between all pairs that have connecting path and excluding all pairs that fall in two different components. So the average path length for this type of network has two different behaviours; when $p_s = 1$, the network has a small average path length and it is bounded, and when $p_s < 1$, the average path length is unbounded. (see *Fig. 3.22*). *Fig. 3.21* shows the path length distribution for **SN** when $p_{sn} = 0.85$, and the path length distribution in the case of fully connected network, the reason why we choose this probability $p_s = 0.85$ is to have bounded path length and be able to compare it with the fully connected case. We intending to use the more satisfactory definition approach for the path length by defining it to be the *harmonic mean* distance between all pairs [6, 98].



(a) (b)

Figure 3.21: (*a*) *Path length distribution for* ***SN*** *with* $p_s = 0.85$, $N = 100$ *nodes and total cell number* $n^2 = 100$ *counted over 5 iterations, with different average path lengths* $6.8, 6.7. 7.5, 8.2, 9.89$, *and* (*b*) *for the fully connected network with the same total cell number, with average path* 6.67.

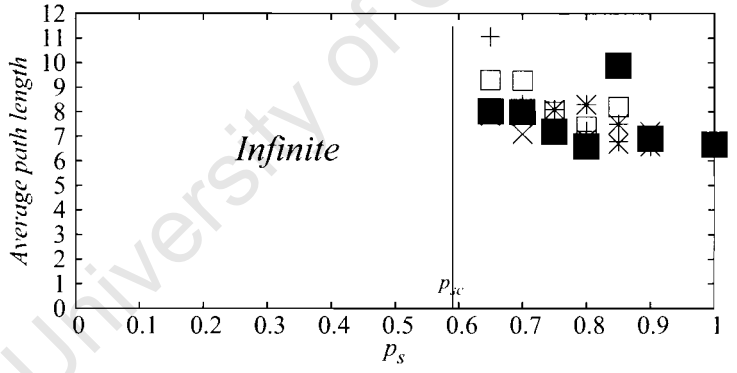


Figure 3.22: *Average path length behaviour for* ***SN*** *below and above the percolation threshold. The dots are the relation between different values of occupation probability and the corresponding path length (we used five different values), also the dots show the different values for the average path length when the probability of a cell being occupied is greater than the percolation threshold.*

3.6 Comparison between SN and RN

So far we have studied some of the statistical properties of **RN** and **SN**, namely, the degree distribution, path length distribution and the cluster size distribution. The expected structure of the two networks varies as the values of p_r and p_s changes. For p_r and p_s below the percolation threshold, p_{rc} and p_{sc}, they consist of small components, the largest cluster in **RN** is a tree and they have a fractal structure for **SN**. For p_r and p_s above the percolation threshold, most of the nodes in these two networks join together in a single giant component. Both **SN** and **RN** exhibit the same type of critical transition at a given threshold, but the values of these threshold are different. $p_{rc} = \frac{1}{(N-1)}$ and $p_{sc} = 0.59$ that is as soon as the number of nodes is more than three or four $p_{rc} \ll p_{sc}$. p_{rc} depends on the number of nodes N, whereas p_{sc} depends only on the lattice type. The maximum degree node for **SN** is always 4, whereas the maximum degree for **RN** is $(N-1)$. The degree distribution in the case of fully connected networks is given by *Eq. 3.17* and in the case when $p_s \neq 1$ the degree distribution for **SN** is given by *Eqs. 3.18* and *3.19*. It has been shown the degree distribution for **RN** is Poisson distribution. The mean degree for **SN** is approximately $4p_s$, whereas the mean degree for **RN** is $p_r(N-1)$. We can find examples of **SN** and **RN** networks which have the same mean degree but the degree distribution around these mean degree will be different. The average path length for **RN** has been shown be vary as $\log(N)$. For **SN** network we were able to calculate analytically the path length distribution only when $p_s = 1$ and we carried out numerical simulation to calculate the path length distribution when $p_s \neq 1$. Further study should be done to estimate the

variation of the average path length of **SN** to compare with $\log(N)$ behaviour of **RN**. The first simulations suggest the average path length for **SN** varies linearly with N.

3.7 Conclusion

In this chapter we have discussed the well known properties of the random networks, such as, degree distribution, cluster size, and the average path length. We have reviewed a work by [6] by introducing a generating function to study the properties of random networks with arbitrary degree distribution. We have also looked at the properties of a 2D spatial lattice network. A comparison between these two types of networks have been done. We have also examined the effect of varying p_r and p_s above and below the percolation threshold p_{rc} and p_{sc} in both networks. Among other things, in the 2D spatial lattice network we have calculate exact theoretical formula for the degree distribution for any p_s and for the average path length when $p_s = 1$. The work presented in this chapter is a tiny fraction of what can be done in this type of spatial network. Further studies would include studying numerically and mathematically the behaviour of the average path length above the threshold and the robustness of such kind of networks. Extending the results obtained on a 2D spatial lattice which are the support of classical percolation models (square lattice with 8 neighbours instead of 4. triangular 2D lattice, cubic 3D lattice, *etc.*). The class of spatial networks that would be obtained when nodes are extracted with different probabilities from this regular lattice as for 2D spatial lattice network that we studied in chapter 3, may exhibit statistical properties which are not only very different

from the classical random networks, but also different from the classical scale-free or small-world networks. They may be used to model spatial data sets in application where links between entities are not randomly distributed but selected according to spatial proximity.

CHAPTER FOUR

Spatial multiscale percolation networkts, (SMPN)

4.1 Introduction

In this chapter we extend our model (discussed in Chapter 3) to generate random networks in a spatial way by considering the case of multiscale structure. We briefly discuss the idea of multiscale percolation system (MPS) and explain how we use it to build two different types of SMPN. We then study three measures of these networks, namely, the degree distribution, the cluster size distribution and the path length distribution.

4.2 Definition of MPS and SMPS

An MPS [99] is obtained as a result of the following iteration procedure. As a basis, we take a given mosaic ($d = 2$) or stacking ($d = 3$) consisting of block of size r_1. These blocks called first-rank blocks, are divided into three classes: $X-, Y-, Z-$ blocks (see *Fig. 4.1*). The choice of the type of block is made randomly, with probability x_1, y_1, z_1respectively

$(x_1 + y_1 + z_1 = 1)$.

In the first step of the construction of the MPS, each first-rank Z-block is divided with preservation of the topology of the original mosaic into n_1^d second-rank blocks of size $r_2 = r_1/n_1$, which are decomposed randomly into $X-, Y-$ and $Z-$ blocks with probabilities x_2, y_2 and z_2 respectively $(x_2+y_2+z_2 = 1)$. In the second step each second-rank $F-$ block is divided analogously into third-rank $X-, Y-$, and $Z-$ blocks of size $r_3 = r_2/n_2$, with probabilities x_3, y_3 and z_3 respectively and so on. In each step the $X-$, and $Y-$ blocks remain in the system and $Z-$ blocks are subject to fractionation into $X-, Y-$ and $Z-$ blocks of smaller size. The parameters n_i are called fractionation multiplicities. So, by repeating of $Z-$ blocks fractionation $N-1$times, we obtain an N-scale percolation system (N-MPS) consisting of $X-$ and $Y-$ blocks of size $r_1, r_2, ..., r_i, ..., r_N$ and $Z-$ blocks of size r_N. So, from the above, the properties of MPS can be determined by the $3N$ parameters

$$r_1, n_i(i = 1, ..., N-1), x_i, y_i(i = 1, ..., N)$$

In the case when the parameters do not depend on the index of the iteration *Fig. 4.2*

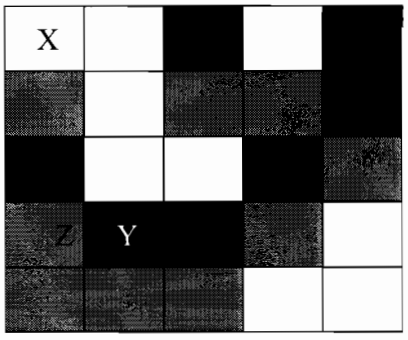


Figure 4.1: *Multiscale Percolation System (MPS)*

i.e. the case of constant probabilities

$$x_i = x, \quad y_i = y, \quad z_i = z = 1 - x - y$$

and constant fractionation multiplicities $n_i = n$, the local structures of the MPS at each step of the construction turn out to be similar (SMPS). As the number of steps tends to infinity, a $Z-$block of rank i can be obtained from a $Z-$ block of rank j as result of a change in all linear scales by a factor of n^{i-j}. In this case all the specific characteristics of the $Z-$ blocks per unit volume are independent of their size.

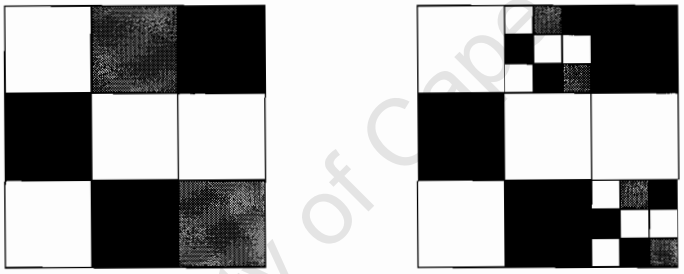


Figure 4.2: *Self-Similar Multiscale percolation System, SMPS.* $d = 2$, $n = 3$, $i = 2$, $x = \frac{4}{9}, y = \frac{3}{9}, z = \frac{2}{9}$.

4.2.1 An Example (Fig.4.3)

In the following example we consider $n = 3$ and $d = 2$, that is a 3×3 grid lattice. We have three possibilities of the cells; black Y cells, white X cells and Z gray cells, which means these gray cells contain again black, white and gray cells. We selected $x = \frac{1}{9}, y = \frac{1}{9}, z = \frac{7}{9}$ in this example see *Fig. 4.3*, we have at the first level $i = 1$; 1 black cell, 1 white cell and the rest are 7 gray cells. Following the structure of SMPS described earlier. At the

second level $i = 2$ we divide each gray cell Z in the lattice to black, white and gray cells, with preservation of the topology of the original lattice. so at each level the number of black cells increases. In this example, at $i = 2$, we expect 7 black cells and when $i = 3$, we expect 7^2 black cells, because we use x, y, z as probabilities, the actual number of black cells varies around the mean value. At the last level, $i_{max} = 4$, the gray cells should be divided again to infinite iterations. But in our finite example, they are replaced by white and black cells in the following proportions $\frac{x}{x+y} = 0.5$ and $\frac{y}{x+y} = 0.5$, in order to obtain the same proportions of white, black areas as would have been obtained if the model was run an infinite.

Figure 4.3: *SMPS structure on* 3×3 *grid,* $x = \frac{1}{9}, y = \frac{1}{9}, z = \frac{7}{9}$, $i = 4$, *where the gray blocks in the last level are replaced by black and white cells in the same proportions as for previous levels.*

4.3 Spatial multiscale percolation network, SMPN

4.3.1 Building networks

We can build several types of networks associated to the same SMPS structure. A first type is denoted by *SMPN1*. An example is shown in *Fig. 4.4*, it is associated to the structure showed in *Fig. 4.3*. The nodes in this type of network are obtained from the black cells at the scale of the last level when the SMPS structure is divided into $(n^2)^{i_{max}}$ cells. where i_{max} is the maximum number of iterations (levels). The number of nodes is always $N < (n^2)^{i_{max}}$. For example, the number of nodes in the network shown in *Fig. 4.4* and associated to the above example *Fig. 4.3* is $N = 3261$. The edges (links) are the neighbouring links. which means that two nodes are linked if the area they represent are neighbours.

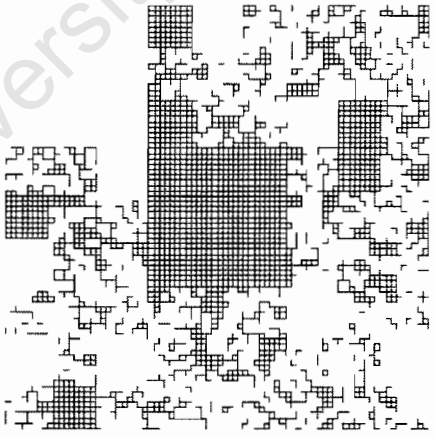


Figure 4.4: *SMPN1 associated to Fig. 4.2*, $N = 3261$

A second type of network is denoted by *SMPN2* and shown in *Fig. 4.5*. This example is associated to the same structure shown in *Fig. 4.3*, the nodes in this type of network are the center of the black cells of different sizes obtained at each level.

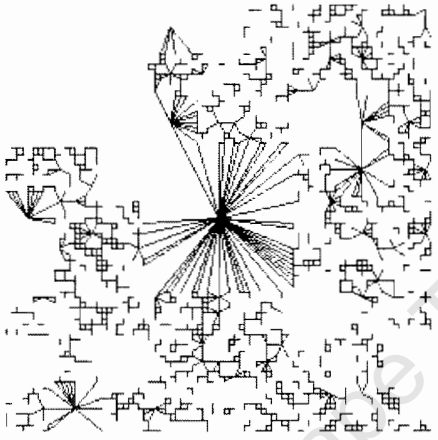


Figure 4.5: *SMPN2 associated to Fig. 4.2*, $N = 1629$

We can calculate the total number of nodes in this type of network. At iteration $i = 1$; the number of black cells (nodes) is yn^2. At $i = 2$, the number of nodes is $yn^2(1 + zn^2)$, and at any step $i - 1$, the number of nodes is $yn^2\left(1 + zn^2 + (zn^2)^2 + ... + (zn^2)^{i-1}\right)$. Summation of this geometric series of ratio zn^2 yields

$$yn^2\left(\frac{(zn^2)^i - 1}{zn^2 - 1}\right).$$

As we said earlier at the last level i_{max} we divide the gray cells into black and white cells according to the respective proportion $\frac{y}{x+y}$ and $\frac{x}{x+y}$, so the number of black cells obtained at the last level is $\frac{y}{x+y}(zn^2)^{i_{max}}$, then the total number of nodes will be

$$N = yn^2\left(\frac{(zn^2)^{i_{max}} - 1}{zn^2 - 1}\right) + \frac{y}{x+y}(zn^2)^{i_{max}}. \tag{4.1}$$

In our example, the exact number of nodes in SMPN2 associated to the above example *Fig. 4.3* is $N = 1629$. So for $x = \frac{1}{9}, y = \frac{1}{9}, z = \frac{7}{9}$ one expect from *Eq. 4.1*, $N = 1614$, which is approximately the same as the exact mean number of black cells. The edges are defined in the same way as of SMPN1, that is, two black cells linked if the cells they represent are neighbours. The number of edges would be difficult to calculate, but in each example, the computer algorithms enables the computing of edges numbers.

4.4 Exploration of the statistical properties of SMPN

4.4.1 Degree distribution

As networks have been often classified by their degree distributions [100], much attention has been paid here to the degree distribution of the SMPN network. We have studied numerically the degree distributions for the two types of networks discussed above, SMPN1 and SMPN2, using the same algorithms defined in chapter 3. We found that there are differences in these two types of networks, that is, they have different degree distribution behaviours. *Fig. 4.6* shows the degree distribution for the SMPN1, the maximum degree value for this type of network is always 4, as 2D spatial lattice network (SN) structure but the proportion of nodes with degree 4 is higher, because of the aggregation structure of the nodes into clusters.

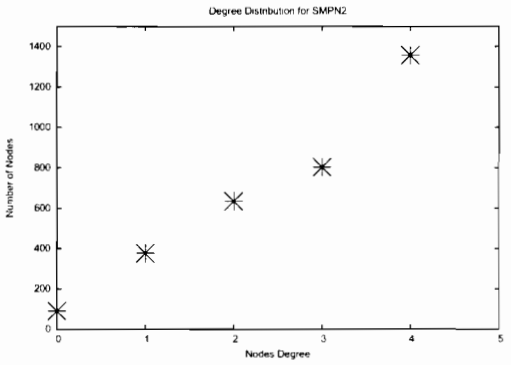


Figure 4.6: *Degree distribution for* ***SMPN1****, with mean degree* 2.18

Fig. 4.7 shows the degree distribution for the SMPN2. As one can see, most of the nodes have few neighbours and a small number of the nodes have a lot of neighbours.

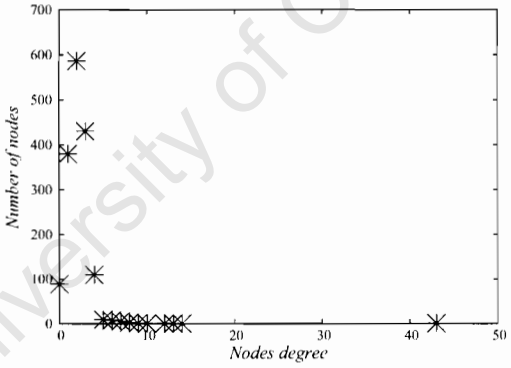
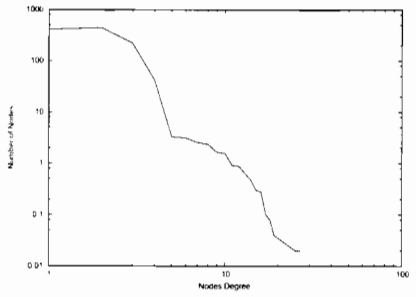
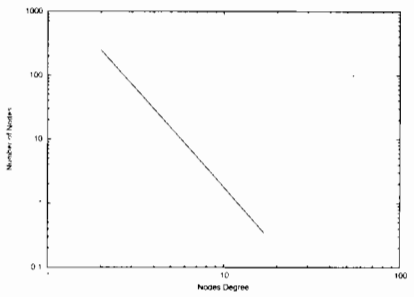


Figure 4.7: *Degree distribution for* ***SMPN2****, with mean degree* 2.91

This is one of the characteristics of the scale- free network and so this kind of network may exhibit a power law degree distribution. *Fig. 4.8(a).* shows the log-log relationship between the nodes degree and the number of nodes in the netwrok (which is not quite linear), its clearly show the characteristic of scale free network which this kind of networks may be close to it. When we gathered the numerical data into three classes of degrees,



(a) The nonlinear Log-Log of the Degree distribution for **SMPN2**.



(b) Log-Log of the three classes of the degree distribution for **SMPN2**.

Figure 4.8: *Degree distribution for* ***SMPN2****.*

the linear regression becomes good *Fig. 4.8(b)*, and we conclude that this type of network may be close to a scale free one.

To check this assumption, we should be able to carry out extended numerical experiments over several levels to a void the boundary problems of the last level iteration, but we could not achieve these numerical experiments because of computer memory limitation. We could also try analytical calculation of the mean theoretical degree distribution. When $i_{max} = 1$ the calculation is the same as in chapter 3, and the result is

$$P(k) = \begin{cases} \frac{(n-2)^2}{n^2}(1-y)^4 + \frac{4(n-2)}{n^2}(1-y)^3 + \frac{4}{n^2}(1-y)^2 & \text{for} \quad k=0 \\ \frac{(n-2)^2}{n^2}4y(1-y)^3 + \frac{4(n-2)}{n^2}3y(1-y)^2 + \frac{4}{n^2}2y(1-y) & \text{for} \quad k=1 \\ \frac{(n-2)^2}{n^2}6y^2(1-y)^2 + \frac{4(n-2)}{n^2}3y^2(1-y) + \frac{4}{n^2}y^2 & \text{for} \quad k=2 \\ \frac{(n-2)^2}{n^2}4y^3(1-y) + \frac{4(n-2)}{n^2}y^3 & \text{for} \quad k=3 \\ \frac{(n-2)^2}{n^2}y^4 & \text{for} \quad k=4 \end{cases} \tag{4.2}$$

To go on calculating the theoretical degree distribution for SMPN when $i_{max} > 1$ it

becomes very difficult when we consider the possible configuration and associated probabilities.

4.4.2 Cluster size distribution

Figs. 4.9 and *4.10* show the cluster size distribution for the SMPN1 and SMPN2 networks. As one can notice, the behaviours of the two networks look almost the same, that is, there is a few larger cluster and a lot of small clusters.

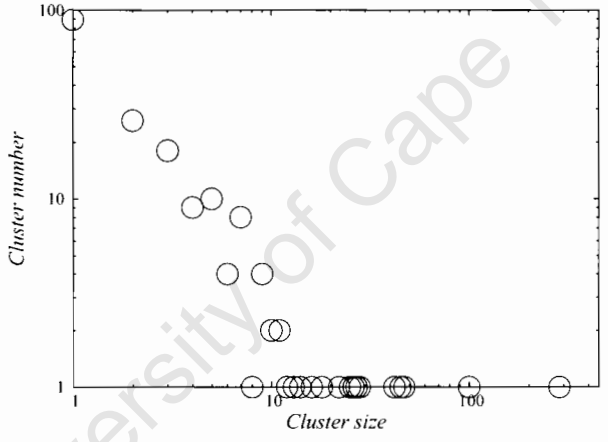


Figure 4.9: *Cluster size distribution for* ***SMPN1***.

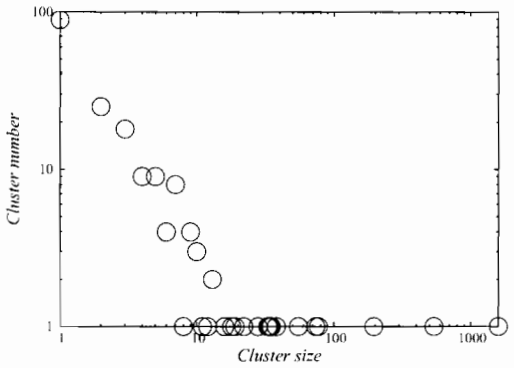


Figure 4.10: *Cluster size distribution for* ***SMPN2****.*

4.4.3 Path length distribution

The numerical calculation of the path length distribution give rather similar results for both SMPN1 and SMPN2. In our 3×3 grids example with $i_{max} = 4$ the maximum path length is close to 100. and the average path length is

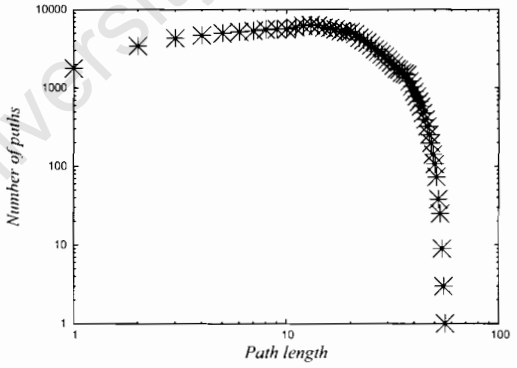


Figure 4.11: *Path length distribution for* ***SMPN1****.*

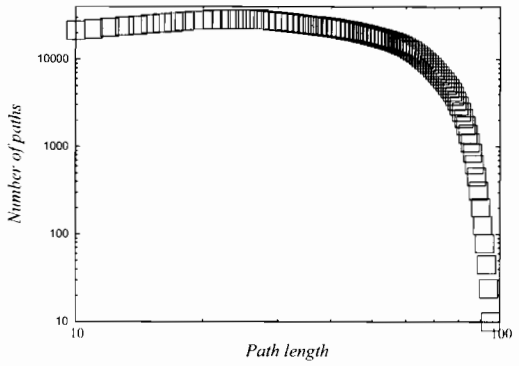
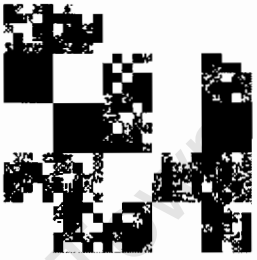


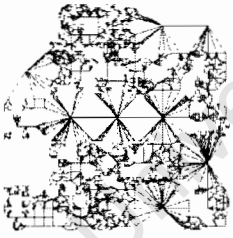
Figure 4.12: *Path length distribution for* ***SMPN2***.

4.5 Other example of SMPN2 networks

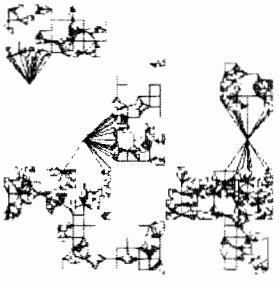
(a) $n^2 = 5 \times 5, N = 2945$, with $x = 0.2. y = 0.4, z = 0.4, i = 3$



(b) $n^2 = 5 \times 5, N = 1750$, with $x = 0.25, y = 0.45. z = 0.3. i = 3$



(c) **SMPN2** associated to (a)



(d) **SMPN2** associated to (b)

Figure 4.13: *Other example of **SMPN2** networks with different probabilities.*

Figure 4.14 below shows the log – log degree distribution for SMPN2.

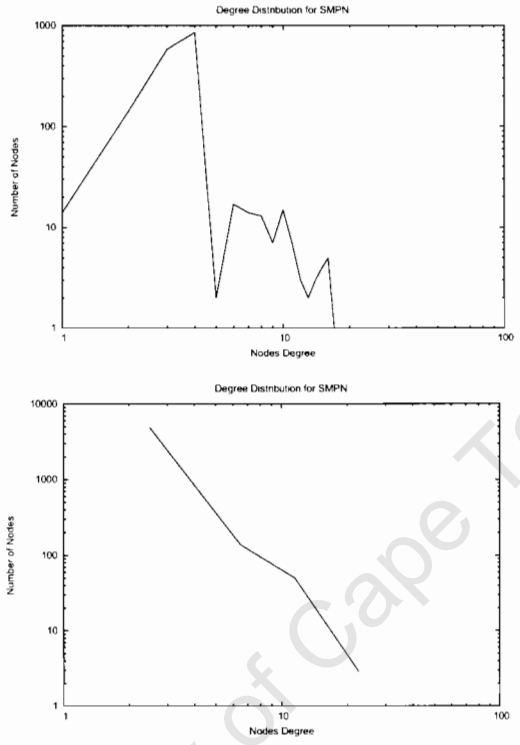


Figure 4.14: *Degree distribution for SMPN2, associated to Fig. 4.13(c)*

4.6 Conclusion

In this chapter we have extended the spatial network model discussed in chapter 3 to multiscale percolation models. We build two different networks, SMPN1 and SMPN2. First network SMPN1 is an extension of the SN percolation model studied in chapter 3 taking into account spatial aggregation of the nodes. Second network SMPN2 is built from the same structure of the multiscale percolation model but defines spatial nodes according to the hierarchy in cell size. We carried out so far only numerical experiment to explore

the statistical properties of these networks. It appears that the SMPN1 exhibits globally the same type of properties that the SN model in chapter 3, but further detailed studies could enable to focus on the difference between 2D percolation lattice with and without nodes aggregation. SMPN2 appears to behave in a way close to scale-free network, with a few nodes having a lot of neighbours and many nodes having only a few neighbours.

In both cases, the definition of the links in these networks according to spatial proximity and to a given geometrical structure leads to a class of spatial network whose properties and use in modelling real world networks may differ strongly from the classical networks and in the network science community. We suggest that SMPN1 could model distribution of species as in chapter 2 taking into account aggregation type 2. While SMPN2 could model road networks and porous media. After the completion of this thesis, we plan to go on exploring the properties of this type of networks, both numerically and analytically.

Chapter Five

Conclusion and further work

This thesis is a first contribution to the analysis of spatial networks by means of statistical properties used so far in the field of network science. Our results are based on a few original numerical exploration and some mathematical calculation which have been presented in chapters 2, 3 and 4. and we rewrite below the conclusion of these different chapters

In chapter 2 we have presented the relationship between the perimeter and abundance, and between perimeter and scale using random patterns. We have done a comparison between two different patterns of distributions, mainly, between the random distribution and aggregated distribution. We have done this by ensuring that all the equations describing the random distribution fit well with the random data (or what ecologists call the 'null hypothesis' or "neutral model") generated by computer simulation, and comparing this case study with two different types of aggregated distribution, type 1 and 2. We have found that there is no difference between aggregated and random distribution which can be detected by using the definition of percolation threshold, when we compare the random distribution of species with the aggregated distribution type 1, but this is not the case when we compare the random distribution with our aggregated distribution, type 2. We

have shown that the relationship between perimeter L and scale a is not the same in the two cases. The underlying power law scaling relating a critical abundance value N_c and the scale a are identical in the case of random distribution and aggregated distribution type 1. We consider that this power law might be an intrinsic property of any spatial data due to mere scaling effects. We can conclude that percolation threshold may not be used to discriminate between different patterns; they are identical for random and aggregated distribution type 1. But it may depend on how the data is aggregated, as shown in our first exploration as regard the aggregated distribution type 2.

Further more in chapter 3 we have discussed the well known properties of the random networks, such as, degree distribution, cluster size, and the average path length. We have reviewed a work by [6] by introducing a generating function to study the properties of random networks with arbitrary degree distribution. We have also looked at the properties of a 2D spatial lattice network. A comparison between these two types of networks have been done. We have also examined the effect of varying p_r and p_s above and below the percolation threshold p_{rc} and p_{sc} in both networks Among other things, in the 2D spatial lattice network we have calculate exact theoretical formula for the degree distribution for any p_s and for the average path length when $p_s = 1$. The work presented in this chapter is a tiny fraction of what can be done in this type of spatial network. Further studies would include studying numerically and mathematically the behaviour of the average path length above the threshold and the robustness of such kind of networks. Extending the results obtained on a 2D spatial lattice which are the support of classical percolation models (square lattice with 8 neighbours instead of 4. triangular 2D lattice, cubic 3D lattice,

etc.). The class of spatial networks that would be obtained when nodes are extracted with different probabilities from this regular lattice as for 2D spatial lattice network that we studied in chapter 3, may exhibit statistical properties which are not only very different from the classical random networks, but also different from the classical scale-free or small-world networks. They may be used to model spatial data sets in application where links between entities are not randomly distributed but selected according to spatial proximity.

Finally in chapter 4 we have extended the spatial network model discussed in chapter 3 to multiscale percolation models. We build two different networks. SMPN1 and SMPN2. First network SMPN1 is an extension of the SN percolation model studied in chapter 3 taking into account spatial aggregation of the nodes. Second network SMPN2 is built from the same structure of the multiscale percolation model but defines spatial nodes according to the hierarchy in cell size. We carried out so far only numerical experiment to explore the statistical properties of these networks. It appears that the SMPN1 exhibits globally the same type of properties that the SN model in chapter 3, but further detailed studies could enable to focus on the difference between 2D percolation lattice with and without nodes aggregation. SMPN2 appears to behave in a way close to scale-free network, with a few nodes having a lot of neighbours and many nodes having only a few neighbours.

In both cases, the definition of the links in these networks according to spatial proximity and to a given geometrical structure leads to a class of spatial network whose properties and use in modelling real world networks may differ strongly from the classical networks and in the network science community. We suggest that SMPN1 could model distribution of species as in chapter 2 taking into account aggregation type 2. While SMPN2 could

model road networks and porous media. After the completion of this thesis, we plan to go on exploring the properties of this type of networks, both numerically and analytically.

As a general conclusion, we feel that further studies should be carried out in at least three directions:

- Carrying out more numerical simulations to explore a broader range of parameter values and modelling assumptions in order to better evaluate the extend of validity of the first results presented here (for example, simulating different types of aggregation in the species abundance model, simulating different types of lattices in the spatial percolation model, analysing many networks with different N, p_r, p_s and different number of iteration levels if any).

- Attempting to obtain, whenever possible, more mathematical results such as, path length distribution in percolation model, scale-free behaviour in chapter 4.

- Analysing real data to see how far they can be modelled by this network presented in this thesis.

Bibliography

[1] M. E. Newman. The structure and function of complex networks. *Society for Industrial and Applied Mathematics Review*, 45, 2003.

[2] Merlot Mathematics Digital Library University of Pittsburgh. An introduction to percolation and many-body physics by computer. *http://fafnir.phyast.pitt.edu/myjava/perc/percTest.html.*

[3] Fangliang He and Stephen P.Hubbell. Percolation theory for the distribution and abundance of species. *PhysRevLett.* 91, 2003.

[4] R. Albert and Albert-Laszlo Barabasi. Statistical mechanics of complex networks. *Rev. Mod. Phys*, 74, 2002.

[5] N.J. Norwood and Albex. Complex network. augmenting the framework for the study of complex systems. *Eur. phys. j., B Cond. matter phys.*, 38, 2004.

[6] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distribution and their applications. *Phys. Rev. E*, 64, 2001.

[7] S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. *Advances in physics*, 51, 2002.

[8] S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks: From biological nets to the internet and www. *Oxford University Press, Oxford*, 2003.

[9] M. E. J. Newman. Models of the small small world. *J. Stat. Phys.* 101, 2000.

[10] B. Hayes. Graphs theory in practice: Part i. *American Scientist.* 88, 2000.

[11] B. Hayes. Graphs theory in practice: Part ii. *American Scientist*, 88, 2000.

[12] Steven H. Strogatz. Exploring complex networks. 2001.

[13] R. Albert, H. Jeong, and A.L. Barabasi. Diameter of the world-wide web. *Nature*, 1999.

[14] A.L. Barabasi, R. Albert, and H. Jeong. Scale-free characteristics of random networks: The topology of the world wide web. *Physsica A*, 281, 2000.

[15] L A. Adamic. The small world web, in lecture notes in computer science. *Springer, New York*, 1696, 1999.

[16] G. W. Flake, S. R. Lawrence, C. L. Giles, and F. M. Coetzee. Self-organization and identification of web communities. *IEEE computer*, 35, 2002.

[17] J. M. Kleinberg, S. R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. The web as a graph: Measurements, model and methods. *in Proceedings of the Interna-*

tional Conference on Combinatorics and Computing, no.1627 in Lecture Notes in Computer Science,pp.1-18, Springer, Berlin, 1999.

[18] A. Broder, R. Kumar, F. Maghoul, P. Raghavan. and S. Rajagopalan. Graph structure in the web. *Computer Networks*, 33, 2000.

[19] A. Vazquez, R. Pastor-Satorras, and A. Vespignani. Internet topology at the router and autonomous system level. *arXiv:cond-mat/0206084*, 2002.

[20] H. Jeong. B. Tombor, R. Albert. ZN Oltvai, and Al Barabsi. The large-scale organization of metabolic networks. *Nature*, 407, 2000.

[21] J. Watts and H. Strogatz. Collective dynamics of "small-world" networks. *NATURE*, 393, 1998.

[22] L. A. N. Amaral, A. Scala, M. Barthlmydagger, and H. E. Stanley. Classes of small-world networks. *PNAS*, 97, 2000.

[23] S. Jespersen, I. M. Sokolov, and A. Blumen. Relaxation properties of small-world networks. *Phys. Rev. E*, 62, 2000.

[24] N. M. Ferguson and G. P. Garnett. More realistic models of sexually transmitted disease transmission dynamics: sexually partership networks, pair models, and moment closure. *Sex. Transm. Dis*, 27, 2000.

[25] S. Gupta, R. M. Anderson, and R. M. May. Netwroks of sexual contacts: Implications for the pattern of spread of hiv. *AIDS*, 3, 1989.

[26] A. S. Klovdahl, J. J. Potterat, D. E. Woodhouse, J. B. Muth, S. Q. Muth, and W. W. Darrow. Social networks and infectious disease: The colorado springs study. *Soc. Sci. Med.* 38, 1994.

[27] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U-. Hwang. Complex networks: Structure and dynamics. *PHYSICS REPORT*. 424, 2006.

[28] Clustering coefficient Wikipedia. http://en.wikipedia.org/wiki/clusteringcoefficient.

[29] Sara Nadiv Soffer and Alexei Vazquez. Clustering coefficient without degree correlation biases. *Phys. rev. E*, 71, 2005.

[30] D. Watts, S. Strogatz, and M. Newman. Networks dynamic, network structure. *www.santafe.edu/research/networkStructure.php*.

[31] Michele Catanzaro, Guido Calderalli, and Luciano Pietronero. Social network growth with assortative mixing. *Physica A*, 338, 2004.

[32] Marian Boguna, Romualdo Pastor-Satorras, and Alessandro Vespignani. Epidemic spreading in complex networks with degree correlations. *arXiv:cond-mat/0301149 v1*, 235, 2003.

[33] R. Albert, H. Jeong, and A.-L. Barabasi. Attack and error tolerance of complex networks. *Nature*, 406, 2000.

[34] J. A. Dunne. R. J. Williams, and N. D. Martinez. Food-web structure and network theory: The role of connectance and size. *Proc. Natl. Acad. Sci. USA*, 99. 2002.

[35] J. A. Dunne. R. J. Williams, and N. D. Martinez. Network structure and biodiversity loss in food webs: Robustness increases with connectance. *Ecology Letters*, 5, 2002.

[36] M. E. J. Newman, S. Forrest, and J. Bathrop. Email networks and the spread of computer viruses. *Phys. Rev. E*, 66, 2002.

[37] R. Solomonoff and A. Rapoport. Connectivity of random nets. *Bulletin of Mathematical Biophysics*, 13, 1951.

[38] A. Rapoport. Contribution to the theory of random and biased nets. *Bulletin of Mathematical Biophysics*, 19, 1957.

[39] P. Erdos and A. Renyi. On random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*. 5. 1960.

[40] B. Bollobas. Random graphs. *Academic Press*, 1985.

[41] E. Almaas, R. V. Kulkarni, and D. Stroud. Characterizing the structure of small-world network. 2002.

[42] Barabasi, Albert-Laszlo, and R. Albert. Scale-free networks., 1999.

[43] Konstantin Klemm and Victor M. Eguiluz. Highly clustered scale-free networks., 2002.

[44] Albert-Laszlo Barabasi and Reka Albert. Emergence of scaling in random networks. *Science*, 286, 1999.

[45] H.-J. Kim and I.-M. Kim. Scale-free network in stock markets. *Journal of the Korean Physical Society*, 40, 2002.

[46] William J. Reed. Forest fires and oil fields as percolation phenomena. *www.math.uvic.ca/faculty/reed/Sci.paper.ps*, 1999.

[47] Harry Kesten. Some highlights of percolation. *ICM*. 1, 2002.

[48] Denis Mollison. Conjecture on the spread of infection in two dimensions disproved. *Nature*. 240, 1972.

[49] Dietrich Stauffer and Amnon Aharony. Introduction to percolation theory., 1994.

[50] Muhammad Sahimi. Application of percolation theory. *Taylor and Francis*, 1994.

[51] Geoffrey Grimmett. Percolation. *Springer*. 321, 1999.

[52] A. Bunde and S. Havlin. Fractals and disordered systems. *Springer Heidelberg*, 1991.

[53] Clustering coefficient Wikipedia. http://en.wikipedia.org/wiki/percolation-theory.

[54] Junqiao Wu. Introduction to percolation theory. *http://www.people.fas.harvard.edu/ wu2/paper1/node1.html*, 1997.

[55] D. Stauffer and A. Aharony. Introduction to percolation theory. *Taylor and Francis, London*. 1994.

[56] Institute for theoretical physics. *http://www.tphys.jku.at/*.

[57] Brain Berkowitz and Isaac Ballberg. Percolation theory and its application to groundwater hydrology. *Water Resources Research*, 29, 1993.

[58] P. R. King, S. V. Buldyrev, N. V. Dokholyan, S. Havlin, Y. Lee, G. Paul, H. E. Stanley, and N. Vandesteeg. Predicting oil recovery using percolation theory. *Petroleum Geoscience*, 7, 2001.

[59] Xu Ruisong. Petroleum and gas research by remote sensing in south china sea. *Symposium on Geospatial Theory, Processing and Applications*, 2002.

[60] E. J. Garboczi, M. F. Thorpe, M. DeVries, and A. R. Day. Universal conductivity curve for a plane containing random holes. *Physical Review A*, 43, 1991.

[61] Antonio Coniglio, H. Eugene Stanley, and W. Klein. Site-bond correlated - percolation problem: A statistical mechanical model of polymer gelation. *Physical Revie Lett.* 42, 1979.

[62] G. Bhat, AH. Hosmani, A. R. Ketkar, and A. R. Paradkar. Percolation theory: application in pharmacy. *Indian Journal of Pharmaceutical Sciences*, 62, 2000.

[63] College of Engineering Northeastern University. http://www1.coe.neu.edu/ emelas/apps.htm.

[64] Brain Berkowitz and Robert P. Ewing. Percolation theory and network modeling application in soil physics. *Geo.Phy*, 19, 1998.

[65] Edith Perrier, Mohammed Altaj, Henri Laurie, and Gareth Witten. Comment on

"percolation theory for the distribution and abundance of species". *Phys. Rev. Lett*, 95, 2005.

[66] J. Brow, D. Mehlman, and G. Stevens. Spatial variation in abundance. *Ecology*, 76, 1995.

[67] Neutral Models in Landscape Ecology: Percolation Theory. http://botit.botany.wisc.edu/courses/460/labs/landperc.html.

[68] William E.Kunin Stephen Hartley and Jack J.Lennon. Scaling down:on the challenge of estimating abundace from occurrence patterns. *The American Naturalist*, 156, 2000.

[69] Fangliang He and Kevin J.Gaston. Estimating species abundance from occurence. *The American Naturalist*, 156, 2000.

[70] Oliver Kruger and George C.McGavin. Macroecology of local inssect communities. *Acta Oecologica*, 21, 1999.

[71] R.H.MacArhur and E.O.Wilson. The theory of island biogeography. *Princcton University Press,Princeton*, 1967.

[72] Charles. J. Krebs. The experimental analysis of distribution and abundance. *Harpers Collin, NY*, 1994.

[73] K. Gaston and T. Blackburn. Pattern and process in macroecology. *Blackwell Scientific*, 2000.

[74] Fangliang He and Kevin J.Gaston. Occupancy-abundance relationships and sampling scales. *Ecography*, 23, 2000.

[75] J. H. Brown. On the relationship between abundance and distribuition of species. *The American Naturalist*, 124, 1984.

[76] K. J Gaston and J.H. Lawton. Effects of scale and habitate on the relationship between regional distribution and local abundance. *Oikos*, 58, 1990.

[77] R. Condit, S. P. Hubbell, J. V. LaFrankie, R. Sukumar, N. Manokaran, R. B. Foster, and P. S. Ashton. Species-area and species individual relationship for tropical trees: a comparison of three 50-ha plots. *Journal of Ecology*, 1996.

[78] J. Green and A. Ostling. Endemic-area relationship: the influence of species dominance and spatial aggregation. *Ecology*, 84, 2003.

[79] Cheryl A.Tosh Belinda Reyers and Albert S.van Jaarsveld. Estimating the abundance of large herbivores in the kruger national park using presence-absence data. *Animal Conservation*, 7, 2004.

[80] D. Wright. Correlations between incidence and abundance are expected by chance. *Journal of Biogeography*, 18, 1991.

[81] John Harte, Erin Conlisk, Annette Ostling, Jessica L. Green, and Adam B. Smith. A theory of spatial structure in ecological communities at multiple spatial scales. *Ecological Monographs*, 72, 2005.

[82] J. Plotkin and H. Muller-Landau. Sampling the species composition of a landscape. *Ecology*, 83, 2000.

[83] Fangliang He Kevin J.Gaston and Jianguo WU. On species-abundance models. *Ecoscience*, 9, 2001.

[84] Daniel S. Falster, Brad R. Murray, and Brendan J. Lepschi. Linking abundance, occupancy and spatial structure: an empirical test of a neural model in an open-forest woody plant community in eastern australia. *Journal of Biogeography*, 28, 2001.

[85] Duncan J. Watts. Six degrres. the science of a connected age. *United Kigdom, William Heiemann*, 2003.

[86] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Principles of statistical mechanics of random networks. *arXiv:cond-mat/0204111*, 2, 2002.

[87] Gary Chartrand. Introductory graph theory. *Boston, Prindle, Weber and Schmidt*, 1997.

[88] M. E. J. Newman. Random graphs as model of networks. *in Handbook of Graphs and Networks, S. Bornholdt and H. G. Schuster (eds.) Wiley-VCH, Berlin*, 2003.

[89] S. Janson, T. Luczak, and A. Rucinski. Random graphs. *John Wiley, New York*, 1999.

[90] Allen W. Wilhite. Department of economics and finance. *http://cas.uah.edu/wilhitea/.*

[91] Albert-Laszlo Barabasi, Rebka Albert, and Hawoong Jeong. Mean-field theroy for scale-free random networks. *Physica A*, 272. 1999.

[92] Agata Fronczak, Piotr Fronzak, and Janusz A. Holyst. How to calculate the main characteristics of random uncorrelated networks. *AIP CONF.PROC.* 776, 2005.

[93] S.N. Dorogovtsev and J. F. F. Mendes. The shortest path to complex networks. *arXiv:cond-mat/0404593*, 4, 2004.

[94] R. Albert. Scale-free networks in cell biology. *Journal of Cell Science*, 118, 2005.

[95] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett.*, 85, 2000.

[96] Agata Fronczak, Piotr Fronczak, and Janusz A. Holyst. Average path length in random networks. *Phys. rev. E*, 70, 2004.

[97] Parongama Sen. Nature of largest cluster size distribution at the percolation threshold. *J. Phys. A: Math. Gen*, 34, 2001.

[98] V. Latora and M. Marchiori. Efficient behavior of small-world networks. *Phys.Rev.Lett*, 87, 2001.

[99] A. V. Neimark. Multiscale percolation systems. *Sov. Phys. JETP.* 69, 1989.

[100] Oskar Hagberg and Carsten Wiuf. Convergence properties of the degree distribution of some growing network models. *Bulletin of Mathematical Biology*. 2006.

APPENDIX A

C code for calculating the perimeter of percolation model considering the internal joins between cells


```
#include<stdio.h>
#include<math.h>
#include<stdlib.h>
#define epsilon  1.e-15

 int  main()
{

  int i ,j,  number =5 , l , h , q, NL = 0;
  double x[number],y[number];
  int  J , jx , jy ;
  int sum = 0,rd ;
  double  cellsize = 50 ,xmax=1000,ymax=500;
  int xcells=xmax/cellsize, ycells=ymax/cellsize;
  int cell_state[xcells][ycells] ;
  int cell_nofdots[xcells][ycells];
  double  prL1 ,prL2,caL,S;
  FILE *data,*order;

  for ( rd = 1; rd <4; rd ++)
    {

      sum  = 0;
      NL   = 0;
  data = fopen("/home/mohammed/wed.dat","w");
  order = fopen("/home/mohammed/order.tex","w");
  fprintf(order,"plot'/home/mohammed/wed.dat'");
```

```
 fclose(order);

 for(i=0;i<number;i++){
   x[i]=(rand()* xmax)/RAND_MAX;
   y[i]=(rand()* ymax)/RAND_MAX;
    printf("%lf %lf\n",x[i],y[i]);
    fprintf(data,"%lf %lf\n",x[i],y[i]);
 }

for (h = 0; h<ycells; h++)
   for (l = 0; l<xcells; l++)

      cell_nofdots[l][h] = 0;
      sum = 0;

for (h = 0; h<ycells; h++){
   for (l = 0; l<xcells; l++)
     {
for (q = 0; q <number; q++)
  if (( x[q] >=  cellsize * l)
            && (x[q] < cellsize *(l+1))
            && ( y[q] >= cellsize * h)
            && (y[q] < cellsize * (h+1))){
     cell_nofdots[l][h] += 1 ;
             sum += 1;
                  }
     }
}
 for (h = 0; h<ycells; h++){
   for (l = 0; l<xcells; l++)
     {
       printf(" %d ",cell_nofdots[l][h]);

     }
   printf("\n");
   }
   printf("\n");

 printf("\n The Total Number of dots are %d \n\n",sum);


 for (h = 0; h < ycells; h++){
   for(l = 0; l < xcells; l++){
```

```
if (cell_nofdots[l][h] != 0)

  /* Replacing the number of dots on each cell by 1 or 0 */
          cell_nofdots[l][h]=1;
cell_state[l][h] = cell_nofdots[l][h];

        printf(" %d",cell_state[l][h]);
    }
    printf("\n");
  }
  printf("\n");

/*Counting The value of the perimeter L  by comparing rows */
  for (i = 0; i<xcells; i++){
    for (j = 0; j<ycells ; j++){
      if ((j == ycells - 1) && cell_state[i][j] != cell_state[i][0] )
 NL+=1;
      if ((j<ycells-1)&&(cell_state[i][j] != cell_state[i][j+1]))
         NL += 1;

    }
  }
    printf("\n The Total Number of dots are %d \n\n",sum);
    printf("\n Perimeter 'L' =  %d  \n",NL);

/*Counting The value of the perimeter L  by comparing columns */
   for (i = 0; i<xcells; i++){
    for (j = 0; j<ycells; j++){
      if ((i == xcells - 1 )&& cell_state[i][j] != cell_state[0][j])
NL += 1;
      if((i<xcells -1 )&& (cell_state[i][j] != cell_state[i+1][j]))
      NL += 1;
    }
   }

   printf("\n  The number of Perimeter joins 'L' =   %d  \n",NL);

/*Counting The Total number of neighbouring joins the map */
  jx = xcells ;
  jy = ycells;
  J = 2 * jx * jy ;

  printf("\n The Total number of neighboring joins a map is %d \n",J);
```

```
 prL1 = 2 * J * cellsize
      * (1 - exp(-(number *cellsize * cellsize )
      /( xmax * ymax)))
      * (exp(-(number *cellsize * cellsize)/(xmax * ymax)));
printf("\n The predicted value  of
                  'L' using p=1-exp(mean) = %lf \n",prL1);

 prL2 = 2 *  J * cellsize
       *( 1 -( pow(( 1 - (( cellsize * cellsize)
       / (xmax * ymax))),number)))
       * (pow (( 1 - ((cellsize * cellsize) / (xmax * ymax))),number));

printf("\n The predicted value of
        'L' for random species using p=1-(1-a/A)^N = %lf \n ",prL2);

caL = NL * cellsize ;

printf("\n The calculated value of 'L' = %lf \n",caL);

 fclose(data);
  system ("gnuplot /home/mohammed/order.tex -");

   }

   return 0;
}
```

C code for building SN networks

```
#include <stdio.h>
#include <math.h>
#include<stdlib.h>
#define dots 150

void gnuplot(const char *gnucommand)
{
  char syscommand[1024];
  sprintf(syscommand, "echo \"%s\" | gnuplot -persist", gnucommand);
  system(syscommand);
}

main()
{

  int i,j,k,rd,h,l,q,sum = 0,temp;
  double x[dots],y[dots];
 double x1[dots],y1[dots];
  double cellsize = 32,xmax=256,ymax=256;
  int xcells=xmax/cellsize, ycells= ymax/cellsize;
  double cells[xcells][ycells];
  int  cell_nofdots[xcells][ycells],xc_size = xcells/cellsize;
  int  yc_size = ycells/cellsize,cell_state[xcells][ycells];
  int  s[dots],label1,label2;

  // FILE *d_dots1 = NULL;
  FILE *c_outt = NULL;
  FILE *c_outt1 = NULL;
  FILE *c_outt2=NULL;
  FILE *Nodes= NULL;
  FILE *command11=NULL;
```

```
  //This will contain commands to gnuplot
  char xi[2500],yi[2500],xj[2500],yj[2500],sti[1000];
  //strings to store commands to gnuplot

  for ( rd = 0; rd <1; rd ++)
    {
      // d_number= fopen("d_number.dat","r");
      c_outt = fopen("con.dat","w");
      c_outt1 = fopen("uncon.dat","w");
      command11=fopen("gnu_command11","w");
      c_outt2=fopen("c_node1.dat","w");
      //fprintf(command11,"plot 'd_number.dat'  \n");
      Nodes= fopen("Nodes8.4.dat","r");
      fprintf(command11,"plot 'Nodes8.4.dat'  \n");
      // Nodes= fopen("cnode.dat","r");
      for(i=0;i<2*dots;i++)
{
  if(i%2==0){
   fscanf(Nodes,"%lf",&x[i/2]);
   // x[i/2]=x1[i/2]*(1+0.5);
   //printf("%lf \n",x[i/2]);
  }
  else
    {
      fscanf(Nodes,"%lf",&y[(i-1)/2]);
      // y[(i-1)/2]=y1[(i-1)/2]*(1+0.5);
    }

}
      for (i=0;i<dots;i++)
{
  sprintf(sti,"%d",i);
  sprintf(xi,"%lf",x[i]);
  sprintf(yi,"%lf",y[i]);
  fprintf(command11,"set label '%s' at %s,%s\n",sti,xi,yi);
  fprintf(command11,"set xrange [0:256] \n");
  fprintf(command11,"set yrange [0:256] \n");
  fprintf(command11,"set xtics 32 \n");
  fprintf(command11,"set ytics 32 \n");
  fprintf(command11,"set grid \n");
}
```

```
for (h = 0; h < ycells; h++)
for (l = 0; l < xcells; l++)
  cell_nofdots[l][h] = 0;
      sum = 0;


      for (h = 0; h < ycells; h++){
for (l = 0; l < xcells; l++)
  {
    for (q = 0; q <dots; q++)
      if (( x[q] >  l*cellsize)
                  && (x[q] < (l+1)*cellsize)
                  && ( y[q] >  h*cellsize)
                  && (y[q] < (h+1)*cellsize)){

cell_nofdots[l][ycells-h-1] += 1 ;
sum += 1;

      }
  }
      }
      for (h = 0; h < ycells; h++){
for (l = 0; l < xcells ; l++)
  {
    printf(" %d ",cell_nofdots[l][h]);

  }
printf("\n");
      }
      printf("\n");

      printf("\n Total number of nodes are %d \n\n",sum);


      for (h = 0; h < ycells; h++)
{
  for(l = 0; l < xcells; l++)
    {
      if (cell_nofdots[l][h] != 0)
cell_nofdots[l][h]=1;
      cell_state[l][h] = cell_nofdots[l][h];
      printf(" %d ",cell_state[l][h]);
    }
```

```
  printf("\n");
}
      printf("\n");


      for (i = 0; i < xcells; i++)
{
  for (j = 0; j < ycells ; j++)
    {

      if ((cell_state[i][ycells-j-1] == 1)
                &&  (j < ycells - 1)
                && (cell_state[i][ycells-j-1] ==
                    cell_state[i][ycells-j-2] ))
{
  for (k=0;k<dots;k++)
    {
      if (x[k] == (i+0.5)*cellsize
                        && y[k] == (j+0.5)*cellsize)
{
  label1 =k;
  break ;
}
    }
  for (k=0;k<dots;k++)
    {
      if (x[k] == (i+0.5)*cellsize
                          && y[k] == (j+1.5)*cellsize)
{
  label2 =k;
  break;
}

    }

  printf("Nodes  %d  and  %d  ARE
                           connected\n",label1,label2);
  fprintf(c_outt2,"%d   %d \n",label1,label2);
  sprintf(xi,"%lf",(i+0.5)*cellsize);
  sprintf(yi,"%lf",(j+0.5)*cellsize);
  sprintf(xj,"%lf",(i+0.5)*cellsize);
  sprintf(yj,"%lf",(j+1.5)*cellsize);
  fprintf(command11,"set arrow from %s,%s
```

```
                              to %s,%s nohead\n",xi,yi,xj,yj) ;
  fprintf(c_outt,"%s %s %s %s\n",xi,yi,xj,yj);

}
    }

}

      for (i = 0; i<xcells; i++)
{
  for (j = 0; j<ycells; j++)
    {

      if ((cell_state[i][ycells-j-1] == 1)
                 &&  ( i < xcells - 1)
                 && (cell_state[i][ycells-j-1] ==
                 cell_state[i+1][ycells-j-1] ))
{
  for (k=0;k<dots;k++)
    {
      if (x[k] == (i+0.5)*cellsize
                          && y[k] == (j+0.5)*cellsize)
{
  label1 =k;
  break ;
}
    }
  for (k=0;k<dots;k++)
    {
      if (x[k] == (i+1.5)*cellsize
                          && y[k] == (j+0.5)*cellsize)
{
  label2 =k;
  break;
}
    }
  printf("Nodes  %d  and  %d  ARE
                          connected\n",label1,label2);
  fprintf(c_outt2,"%d  %d \n",label1,label2);
  sprintf(xi,"%lf",(i+0.5)*cellsize);
  sprintf(yi,"%lf",(j+0.5)*cellsize);
  sprintf(xj,"%lf",(i+1.5)*cellsize);
  sprintf(yj,"%lf",(j+0.5)*cellsize);
```

```
  fprintf(command11,"set arrow from %s,%s
                          to %s,%s nohead\n",xi,yi,xj,yj) ;
  fprintf(c_outt,"%s %s %s %s\n",xi,yi,xj,yj);

}
    }
}

      fprintf(command11,"replot\n");
      fclose(command11);
      fclose(c_outt);
      fclose(c_outt1);
      fclose(c_outt2);
      //fclose(d_dots1);
      fclose(Nodes);
      system("gnuplot gnu_command11 -\n");
    }
}
```

The Python and C codes for calculating the statistical properties of networks

C.1 C code gives a list of connected nodes

```
#include <stdio.h>
#include <math.h>
#define dots 20

main()
{

  int i,j;
  int x[500],y[500],dfs[500][500];
  int n,r;
  FILE *node=NULL;
  FILE *c_size ;
  node= fopen("c_node1.dat","r");
  c_size = fopen("clus_size.dat","w");
  //connect=fopen("cout2.dat","r");

  i=0;
  while(!feof(node))
    {
      fscanf(node,"%d",&x[i]);
      fscanf(node,"%d\n",&y[i]);
        printf("x[%d]=%d  y[%d]=%d\n",i,x[i],i,y[i]);

      i=i+1;
    }
```

```
int k=i;

  for(i=0;i<dots;i++)
    {
     for (j=0;j<dots;j++)
{
  dfs[i][j]=-1;
}
    }

  j=1;r=1;
  dfs[j][r]=x[0];
  dfs[j][r+1]=y[0];
  printf("%d %d ", dfs[j][r],dfs[j][r+1]);
  fprintf(c_size,"%d %d ",dfs[j][r],dfs[j][r+1]);

  for(i=0;i<k-1;i++)

    {

  if(x[i+1]==x[i])

    {

    dfs[j][r+2]=y[i+1];
    printf("%d ",dfs[j][r+2]);
    fprintf(c_size,"%d ",dfs[j][r+2]);
    r++;

   }

  else

    {

   j++;
   r=1;
   dfs[j][r]=x[i+1];
   dfs[j][r+1]=y[i+1];
   printf("\n");
   fprintf(c_size,"\n");
   printf("%d %d ", dfs[j][r],dfs[j][r+1]);
   fprintf(c_size,"%d %d ",dfs[j][r],dfs[j][r+1] );
```

```
      }

  }

   printf("\n");
  fprintf(c_size,"\n");
  fclose(c_size);

}
```

C.2 The Python codes for calculating the cluster size distribution for networks

```
import StringIO
import sys
file_iin = open('result2.dat','r')
out_file = sys.stdout
file_out3 = open('result3.dat','w')
lines = [map(int,line.split()) for line in file_iin] # make lines as int lists

def length_compare(a, b):
    if len(a) == len(b):
        return 0
    elif len(a) < len(b):
        return 1
    else:
        return -1

lines.sort(cmp = length_compare)
#print lines

list_num = []
o=[]

for i1 in range (len(lines)):

    include = True
    for i2 in range(i1):
        if i1 == i2:
            pass
```

```
        else:
            if len(lines[i1]) > len(lines[i2]):
                for c in lines[i2]:
                    if c in lines[i1]:
                        include = False
            else:
                for c in lines[i1]:
                    if c in lines[i2]:
                        include = False
    if include:
        list_num.append(lines[i1])
#    if not (include):
  #     list_num.append(lines[i1])
#print list_num

for line in list_num :
    list_num = []
    for i in range (len(line)-1):
        list_num.append(line[i])

    list_num.sort()
    x = 0

    while x < len (list_num) - 1 :
        if list_num[x] == list_num[x+1] :
            del list_num[x+1]
        else:
            x = x + 1
    print list_num
    print "Cluster of size : ",len(list_num)


file_out3.close()
file_iin.close()

res_file=open('result2.dat','r')
mlf = res_file.readlines()
list_num = []
o=[]

for i1 in range (len(mlf)):

    include = True
```

```
    for i2 in range(i1):
        if i1 == i2:
            pass
        else:
            if len(mlf[i1]) >= len(mlf[i2]):
                for c in mlf[i2]:
                    if c in mlf[i1]:
                        include = False
            else:
                for c in mlf[i1]:
                    if c in mlf[i2]:
                        include = False
    if include:
        list_num.append(mlf[i1])
 #   if not (include):
    #    list_num.append(mlf[i1])
#print list_num

for line in list_num :
    list_num = []
    for i in range (len(line)-1):
        list_num.append(int(line[i]))

    list_num.sort()
    x = 0

    while x < len (list_num) - 1 :
        if list_num[x] == list_num[x+1] :
            del list_num[x+1]
        else:
            x = x + 1
    print list_num
    print "Cluster of size : ",len(list_num)



res_file.close()


#import StringIO
import sys
file_in = open('clus_size.dat','r')
out_file = sys.stdout
```

```
file_out1 = open('result1.dat','w')
lines = [map(int,line.split()) for line in file_in] # make lines as int lists
for i, line in enumerate(lines):
    out = []
    for digit in line:
        out.append(digit)
      #  print 'out',out
        for followingline in lines[i+1:]:
            if digit in followingline:
                out.extend([x for x in followingline if x != digit])
    out_file.write(' '.join(map(str, out))+"\n")
    file_out1.write(' '.join(map(str, out))+"\n")
file_in.close()
file_out1.close()
```