NATIONAL & KAPODISTRIAN UNIVERSITY OF ATHENS
DEPARTMENT OF SCIENCES
POSTGRADUATE PROGRAM IN LOGIC, ALGORITHMS AND COMPUTATION

$\mu\Pi\lambda\forall$

MASTER THESIS

Percolation on Small-World Networks

Elena Chatzigeorgaki
A.M. 200504

Supervisor: Evangelos Kranakis

Athens
May 2010

Many Thanks to...

First of all I would like to thank professors Y. Moschovakis and C. Dimitrakopoulos for accepting me in the postgradute program in Mathematical Logic and Complexity and through their courses helped me broaden my horizons of knowledge.

I truly thank my supervisor Evangelos Kranakis for this Master Thesis for his valuable assistance and patience.

I would also like to thank professors V. Zisimopoulos, D. Thilikos, Y.Emiris, E. Koutsoupias, M. Koumparakis, E. Raptis, E. Floratos for their interesting courses and their ability to transfer knowledge.

Many thanks also to my fellow students Maria Strataki, Katerina Petsi, Hlias Rokos, Pyrros Chaidos, Kostas Sdrakas, Gregory Galiatsatos, William Billy Karageorgos and Antonis Varvitsiotis for making the years of my studies a pleasant ride.

Finally, I would like to thank professor Konstantin Halatsis because he always believed in me. Most of all I thank Professor Yannis Ioannidis for his patience, kindness and guidance who helped me get through the most rough period of my life.

Contents

List of Figures

1 Introduction

A network is a set of items (vertices) connected by edges. In the real world many systems take the form of network such as the World Wide Web, the internet, social networks of acquaintances, networks of citations between papers, neural networks, metabolic networks, food webs etc. These networks exhibit a number of statistical properties that we have to study in order to understand them. First we have to define the properties that characterize the structure of networks and then create models of networks that can help us understand the meaning of these properties. Many network models have been proposed but we will extensively analyze the small world model which is based one the "small-world phenomenon" - the principle that we are all linked by a short chain of acquaintances as was proved by the studies of Stanley Milgram in the 1960's. After exploiting the advantages of the small world model we will study the epidemic behavior in such a model. This master thesis is structured as follows: In section 2 we give an introduction to site and bond percolation and give an example of how percolation works on 1 and 2 dimensional lattices. In section 3 we give a small introduction to some theoretic preliminaries along with the most basic small world models proposed so far. In section 4 we extensively study the site and bond percolation problem on a Newman-Watts small world model and finally on section 5 we study the greedy routing problem on the r-Harmonic small world model.

2

What Is Percolation?

2.1 General Introduction

Percolation theory is a field mostly studied by physicists but covers a wide range of applications useful in other sciences like chemistry, mathematics and materials science. It was introduced to answer questions like:

- If we put a porous rock underwater, will the water reach it's center?
- How far from each other should we plant trees in an orchard (forest) in order to minimize the damages from a fire outburst?
- How fast will an infectious disease spread? How long before it causes a pandemic?

One of its most popular applications is the behavior of fluids in porous media, mostly used to improve the productivity of natural gas and oil wells. In physics, percolation theory is used to study the flow of electricity in two dimensional random resistor networks. Percolation models are used in biology to study evolution and also in social sciences to study phenomena like how fast a rumor can spread.

In general, percolation is used to study dynamical systems and thus considered a branch of statistical mechanics. Percolation systems go through phase transitions particularly around a critical point or threshold. What is most interesting about percolation theory is that it provides a simple model of random media yet realistic towards each application.

2.2 Sites And Lattices

Consider a square lattice, i.e. an infinite array of squares, denoted by $\mathbb{Z}^2$. (Figure 2.1(a)[1]). A fraction of squares are filled with a dot in the center, while the other squares are left empty as in Figure 2.1(b). We now define a *cluster* as a group of neighbor squares occupied with these dots. These clusters are encircled in Figure 2.1(c).

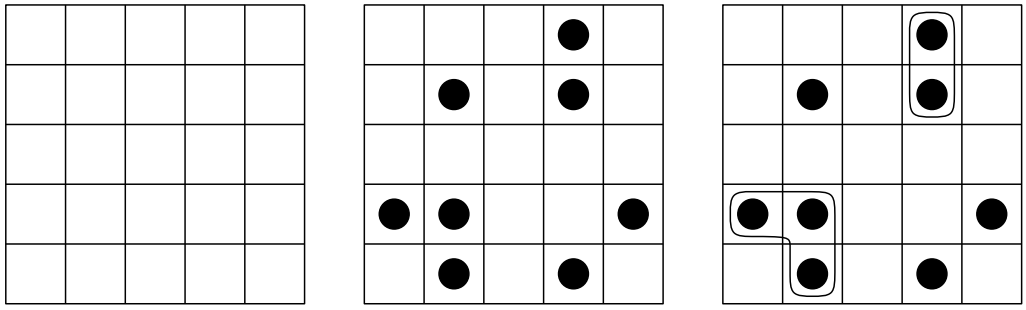


Figure 2.1: Definition of percolation and its clusters

Squares are called *neighbors* (or *nearest neighbor sites*) if they have one side in common but not if they only touch one corner (*next nearest neighbors*). (Figure 2.2)

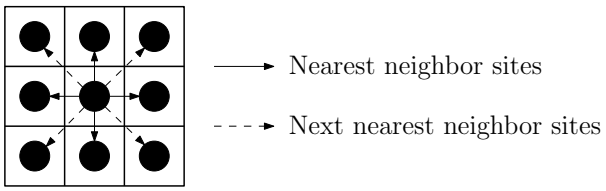


Figure 2.2: Neighbors of a site in a square lattice

All sites within one cluster are connected to each other by one unbroken chain of nearest neighbor links from one occupied square to an occupied neighboring square. *Percolation theory* deals with the number and properties of these clusters. The first question arising is: how are these dots distributed on the square lattice? The answer is that the occupation status of any square on the lattice is *independent* of the occupation status of the its neighbors, i.e. each square is randomly, independently, occupied with a probability p, $0 \leq p \leq 1$. That means, if we have N squares (with N being a very large number) then the expected number of occupied squares is pN and the expected number of *unoccupied* or *empty* squares are the remaining $(1-p)N$.

We concentrate here with the case of *random percolation*. Each site of a very large lattice is occupied randomly with probability p independent of its neighbors. Percolation theory deals with the clusters thus formed.

[1] For obvious reasons the following figures will be of finite dimensions.

In Figure 2.3[2] we see an example of how percolation works on a two dimensional lattice. For small values of the occupation probability ($p = 0.15, 0.30, 0.45$) some disconnected parts are distinguished in the lattice, but for $p \geq 0.59$ we can see that one cluster extends from top to bottom and from left to right of the lattice without intermediate gaps. We say that this cluster *percolates* through the system. Near that concentration p_c, where for the first time a cluster is formed, a lot of peculiar phenomena are observed. These aspects are called *critical phenomena* and the theory attempting to describe them is the *scaling theory*.

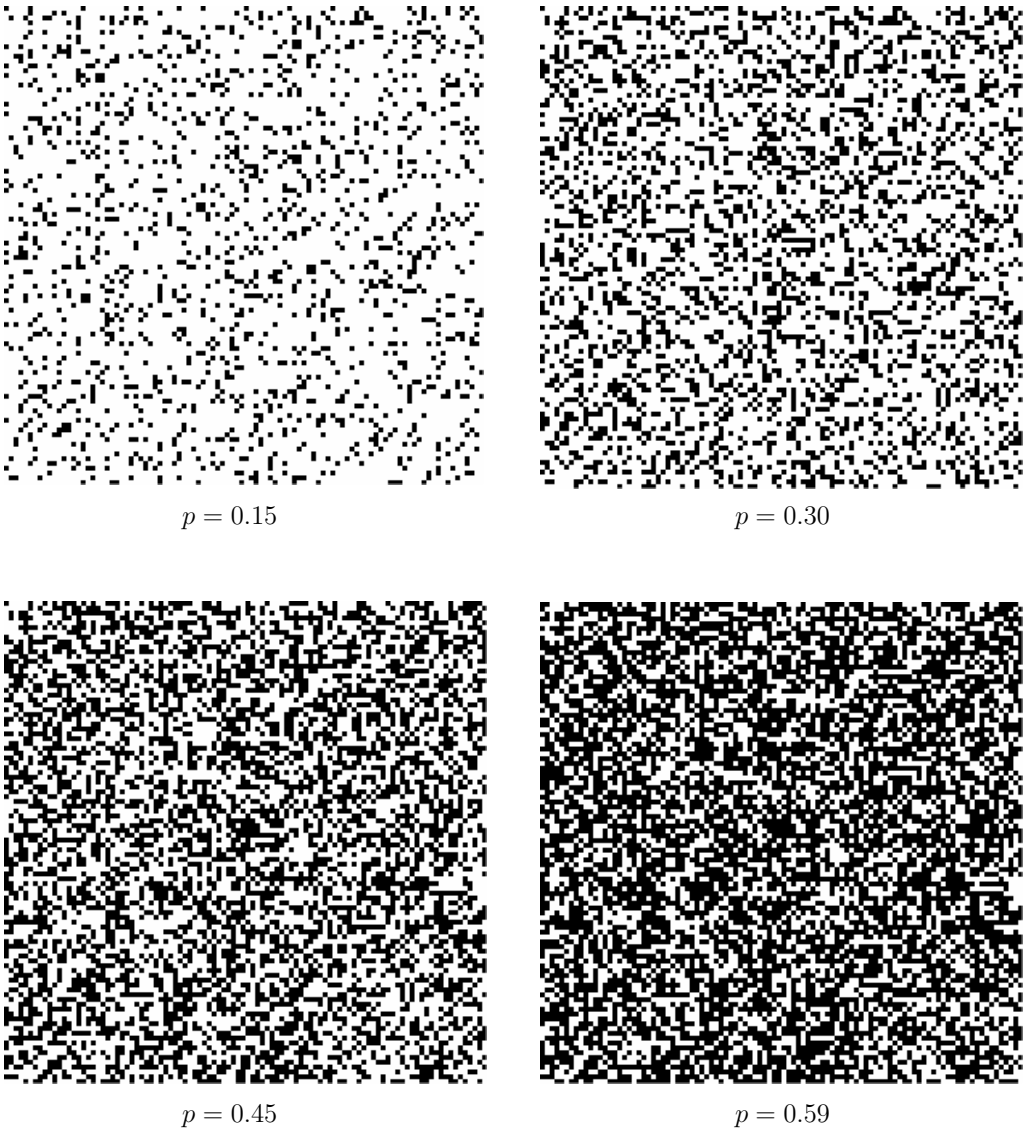
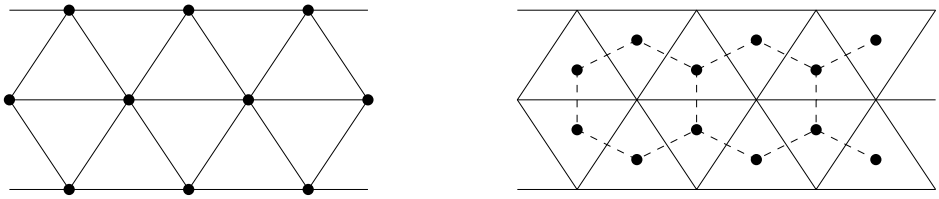


Figure 2.3: Example of site percolation

[2]Example generated with MATLAB©

2.3 Lattices

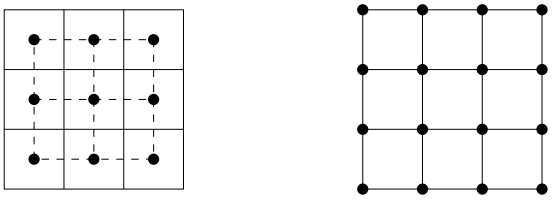
So far we've seen examples of percolation only on square lattices but in reality there are many different lattices or other two or three dimensional structures (graphs in general) upon which we study percolation phenomena. In two dimensions we also have the *triangular* lattice (Figure 2.4(a)) where every intersection of lines is a lattice site and the *honeycomb* (or *hexagonal*) lattice where the centers of the triangles are lattice sites (Figure 2.4(b)).



(a) Triangular Lattice (b) Honeycomb Lattice

Figure 2.4: Other lattices in two dimensions

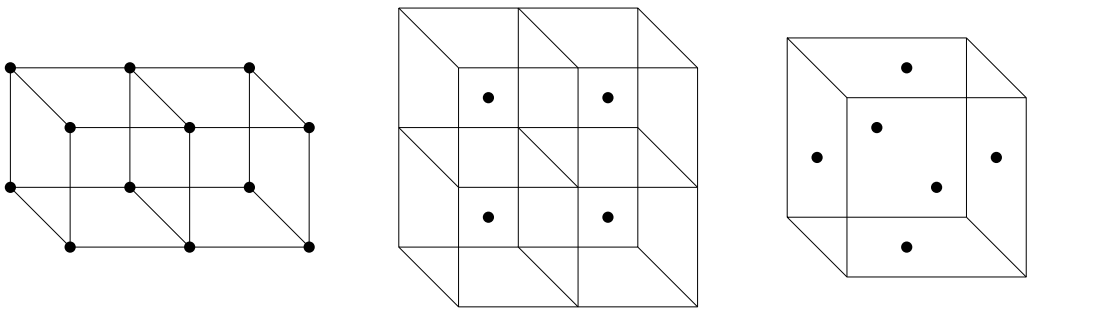
We defined the square lattice through the centers of the squares (Figure 2.5(a)). We could have also defined it equivalently through the points where the lines cross (square corners) as in Figure 2.5(b).



(a) Square Centers (b) Square Corners

Figure 2.5: Square Lattices

In three dimensions we have the *simple cubic lattice* (Figure 2.6(a)), the *body centered cubic (BCC)* lattice (Figure 2.6(b)), the *face-centered cubic (FCC)* lattice (Figure 2.6(c)), the diamond lattice and others. For dimensions higher than 3 we study the hypercubic lattice.



(a) Simple Cubic Lattice (b) BCC: Body-Centered Cubic Lattice (c) FCC: Faced-Centered Cubic Lattice

Figure 2.6: Cubic Lattices

2.4 Site Percolation, Bond Percolation

For all the aforementioned lattices, each site is randomly and independently occupied with probability $p, 0 \leq p \leq 1$ and empty (unoccupied) with probability $(1-p)$. Clusters are thus formed as groups of neighboring occupied sites. So far we've defined *site percolation*. Its counterpart is called *bond percolation* and its defined as follows. In bond percolation every lattice site is occupied. Each *line* can be an *open bond* with probability $p, 0 \leq p \leq 1$ or a *closed bond* with probability $(1-p)$. A cluster is then a group of sites connected by open bonds (Figure 2.7).

It has to be noticed that when measuring the *size* of a cluster, one has to define whether one counts the *site content* or the *bond content*. For example the 3rd encircled cluster in Figure 2.7 consists of two occupied sites connected with an open bond to each other and with closed bonds to all other neighboring sites. This is called a cluster of size two in site percolation but it is called a cluster of size one in bond percolation.

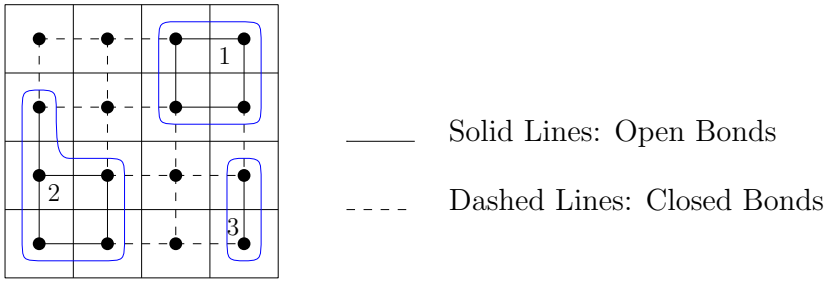


Figure 2.7: Bond percolation on the square lattice

2.5 Percolation Threshold

The *percolation threshold* p_c is the critical fraction of lattice squares that must be occupied in order to create a continuous path of nearest neighbors from one side of the lattice to the opposite side[3].

- For all $p > p_c$ there is a cluster extending from one side of the system to the other, whereas
- for $p < p_c$ no such infinite cluster exists.

Computer simulations do not allow infinite computations, so essentially this is an asymptotic value. Any effective threshold values obtained numerically or experimentally have to be carefully *extrapolated to infinite system size*. The ideal case is when one has a mathematically exact calculation for p_c where no such extrapolation is needed. Mathematical methods to calculate the exact percolation threshold are restricted to at most two dimensions because of our experience in the field of *phase transition* where 3-dimensional problems in general cannot be solved exactly. The review of Essam (1972), as well as Kesten (1982) explain how 2-dimensional thresholds can be derived mathematically for many simple lattices. Progress is not easy in this field. For the *square bond percolation* problem, it took about two decades from the first numerical estimates in 1960 to a mathematical proof that

[3]In random graphs this process is called the *emergence of a giant component* as we will see later in this text.

yield the exact threshold $p_c = \frac{1}{2}$. We also know $p_c = \frac{1}{2}$ for the *triangular site percolation*, $p_c = 2\sin\frac{\pi}{18}$ for the *triangular bond percolation*, and $p_c = 1 - 2\sin\frac{\pi}{18}$ for the *honeycomb bond percolation* problem (Table 2.1).

Lattice	Site Percolation	Bond Percolation
Honeycomb	0,6962	0,65271
Square	0,592746	$\frac{1}{2}$
Triangular	$\frac{1}{2}$	0,34729
Diamond	0,43	0,388
Simple Cubic	0,3116	0,2488
BCC	0,246	0,1803
FCC	0,198	0,119
d=4 Hypercubic	0,197	0,1601
d=5 Hypercubic	0,141	0,1182
d=6 Hypercubic	0,107	0,0942
d=7 Hypercubic	0,089	0,0787

Table 2.1: Site and bond percolation thresholds in different lattices

In all of the above examples clusters are defined as groups of nearest neighbors which are occupied or connected with open bonds. One may allow next-nearest neighbors to form clusters, so in the square lattice site percolation problem, diagonally occupied sites may also form clusters. One can also add long range contacts (or shortcuts) in a lattice and then study percolation phenomena. In the latter case, percolation thresholds tend to zero if the connection range goes to infinity. One may even get rid of the lattice and look at circles distributed randomly on a piece of paper! Finally percolation phenomena can be studied in many types of graphs as is the case with small-world networks where in an initial ring lattice structure, long range contacts are introduced according to some probabilistic experiment.

2.6 Exact Solution In One Dimension

The percolation problem in one dimension can be solved exactly and some aspects of that solution seem to be valid for higher dimensions.

Consider an infinite long chain where lattice sites are placed in fixed distances (Figure 2.8). Each site is occupied with probability p. A cluster is thus formed by successive occupied sites that have no empty site between them. To separate one cluster from the other clusters formed in the lattice, the left and right end neighbors of the cluster must be empty sites. As shown in Figure 2.8, the central cluster consists of five occupied sites and the left and right neighbor sites of this cluster are empty sites. As mentioned earlier each site is occupied with probability p, thus the probability of a site being empty is $(1 - p)$.

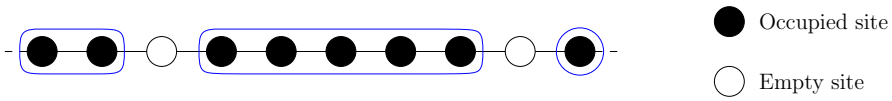


Figure 2.8: Percolation clusters in an one dimensional lattice

Since all sites are occupied randomly, and random percolation consists of statistically independent events, the probability[4] of two arbitrary sites being occupied is p^2, for 5 being occupied is p^5 and for s sites being occupied the probability is p^s. The probability of having an empty neighboring site is $(1-p)$ and the events that the two ends of a cluster are empty are statistically independent, therefore the total probability that a fixed lattice site is the left end of a 5-cluster is $p^5(1-p)^2$.

The next step is to calculate the number of 5-clusters in the whole chain. If the total length of the chain is L, with $L \to \infty$, much larger than the cluster length, then the total number of 5-clusters - if we ignore the small number of sites at the end of the chain for which there is no place for 5 occupied and 2 empty sites - is $Lp^5(1-p)^2$. From now on it is practical to talk about the number of clusters per lattice site, which is:

$$\frac{\text{Total number of 5-clusters}}{\text{Lattice sites}} = \frac{L \cdot p^5(1-p)^2}{L} = p^5(1-p)^2$$

This number is independent of the lattice size L and equals the probability that a fixed site is the end of a 5-cluster. We can also generalize this number in the case of clusters of size s. We define n_s to be the number of clusters of size s per lattice site:

$$n_s = p^s(1-p)^2 \tag{2.1}$$

This normalized *cluster number* equals the probability in an infinite chain, of an arbitrary site being the left *end* of the cluster. The probability that an arbitrary site is *part* (and not only the left end) of an s-cluster is $n_s s$, because now that site can be any of the s sites of the cluster. Moreover a single occupied site with two empty neighbors is a cluster of size unity. Thus every occupied site in the chain belongs to a cluster. The probability that an arbitrary occupied site belongs to any cluster, is equal to the probability p that it is occupied, i.e.:

$$\sum_{s=1}^{\infty} n_s s = p \quad (p < p_c) \tag{2.2}$$

This can be also verified using a trick to calculate a sum by expressing it as a derivative

$$\frac{d}{dx}\sum_{i=1}^{\infty} f_i(x) = \sum_{i=1}^{\infty} \frac{d}{dx} f_i(x), \tag{2.3}$$

by the following simple proof:

[4] Using the product property of probabilities of independent events

$$\sum_{s=1}^{\infty} n_s s \overset{(2.1)}{=} \sum_{s=1}^{\infty} p^s (1-p)^2 s = (1-p)^2 \sum_{s=1}^{\infty} p^s s = (1-p)^2 \sum_{s=1}^{\infty} p \frac{d(p^s)}{dp} \overset{(2.3)}{=} (1-p)^2 p \frac{d(\sum_{s=1}^{\infty} p^s)}{dp}$$
$$= (1-p)^2 p \frac{d(\sum_{s=0}^{\infty} p^s - 1)}{dp} = (1-p)^2 p \frac{d(\frac{1}{1-p})}{dp} = (1-p)^2 p \frac{1}{(1-p)^2} = p$$

The last step is to calculate the percolation threshold. For $p = 1$ every site is occupied forming a cluster of size L. For $p < 1$ a chain of length L will have on average $(1-p)L$ empty sites. As $L \to \infty$ at fixed p, $(1-p)L$ also tends to ∞. Thus there will be at least one empty site somewhere in the chain breaking the sequence of continuous occupied sites. In other words there is no one-dimensional percolating cluster for $p < 1$. Therefore the percolation threshold is unity.

$$p_c = 1$$

2.7 Average Cluster Size

So far we know that the probability that an arbitrary site (occupied or not) belongs to a cluster of size s is $n_s s$ and the probability that an arbitrary site belongs to any finite cluster is $\sum_{s=1}^{\infty} n_s s$. Therefore the probability that the cluster to which an occupied site belongs contains exactly s sites is:

$$w_s = \frac{n_s s}{\sum_{s=1}^{\infty} n_s s} \tag{2.4}$$

Now we can define the *mean cluster size* S as the probability of hitting some cluster *site*. We can calculate the mean cluster size S explicitly:

$$S = \sum_{s=1}^{\infty} w_s s \overset{(2.4)}{=} \sum_{s=1}^{\infty} \frac{n_s s^2}{\sum_{s=1}^{\infty} n_s s} cc \sum_{s=1}^{\infty} \frac{n_s s^2}{p} \overset{(2.1)}{=} \sum_{s=1}^{\infty} \frac{p^s (1-p)^2 s^2}{p}$$
$$= \frac{(1-p)^2}{p} \sum_{s=1}^{\infty} s^2 p^s = \frac{(1-p)^2}{p} \left(\sum_{s=1}^{\infty} s^2 p^s - s p^s + \sum_{s=1}^{\infty} s p^s \right)$$
$$= \frac{(1-p)^2}{p} \left(p^2 \sum_{s=1}^{\infty} s(s-1) p^{s-2} + p \sum_{s=1}^{\infty} s p^{s-1} \right) = \frac{(1-p)^2}{p} \left(p^2 \sum_{s=1}^{\infty} \frac{d^2(p^s)}{dp^2} + p \sum_{s=1}^{\infty} \frac{d(p^s)}{dp} \right)$$
$$= \frac{(1-p)^2}{p} \left(p^2 \frac{d^2(\sum_{s=1}^{\infty} p^s)}{dp^2} + p \frac{d(\sum_{s=1}^{\infty} p^s)}{dp} \right) = \frac{(1-p)^2}{p} \left(p^2 \frac{d^2(\sum_{s=0}^{\infty} p^s - 1)}{dp^2} + p \frac{d(\sum_{s=0}^{\infty} p^s - 1)}{dp} \right)$$
$$= \frac{(1-p)^2}{p} \left(p^2 \frac{d^2(\frac{1}{1-p})}{dp^2} + p \frac{d(\frac{1}{1-p})}{dp} \right) = \frac{(1-p)^2}{p} \left(p^2 \frac{d(\frac{1}{(1-p)^2}}{dp^2} + p \frac{1}{(1-p)^2} \right)$$
$$= \frac{(1-p)^2}{p} \left(p^2 \frac{2}{(1-p)^3} + p \frac{1}{(1-p)^2} \right) = \frac{2p}{1-p} + 1 = \frac{1+p}{1-p}, \quad (p < p_c) \tag{2.5}$$

The mean cluster size diverges as we approach the percolation threshold. If there exists an infinite cluster above the threshold, then slightly below there exist very large (finite) clusters. This implies that slightly below the threshold, a suitable average over these clusters is also getting very large.

2.8 Percolation In 2 Dimensions

Calculating the exact percolation threshold in one dimension was quite an easy task, but we cannot apply the same principles in higher dimensions. Consider a square lattice as in Figure 2.9. The probability that an arbitrary occupied site is a cluster of size 1 is: $n_1 = p(1-p)^4$, where p is the probability of the site being occupied, $(1-p)^4$ is the probability that its four nearest neighbors are empty and the occupation status of these five sites happens independently. We can easily calculate the average number of clusters of size 2 per lattice site. That is: $n_2 = 2p^2(1-p)^6$, where p^2 is the probability of two sites being occupied, $(1-p)^6$ is the probability their six nearest neighbors being empty, the occupation status of these sites happens independently and the pair can be oriented either *horizontally* or *vertically*, i.e. we have two *configurations* of a cluster of size 2 (Figure 2.9). For higher cluster sizes it is not easy to calculate their average number and that is because there are plenty of cluster configurations (different shapes and various rotations) called *lattice animals*[5]. For example in Figure 2.10 there is a list of the 19 cluster configurations on the square lattice for $s = 4$ along with their correspondent probabilities.

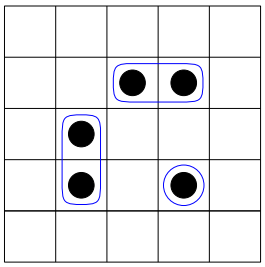


Figure 2.9: Clusters of sizes 1 and 2

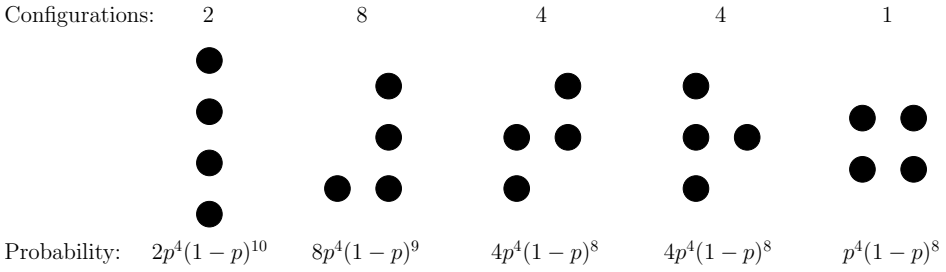


Figure 2.10: Cluster configurations and probabilities for $s = 4$

Thus the average number of clusters of size 4 is:

$$n_4 = 2p^4(1-p)^{10} + 8p^4(1-p)^9 + 4p^4(1-p)^8 + 4p^4(1-p)^8 + p^4(1-p)^8$$

For $s = 5$ there exist 63 configurations and up to $s = 24$, on the square lattice, there are approximately 10^{13} configurations so it is not effective to count these cluster animals. Instead, we classify them according to the number of empty neighbors each of them has. The number of empty neighbors of a cluster, denoted by t, is called its *perimeter*. The number of lattice animals with size s and perimeter t is denoted by g_{st}. Now we can express the average number of clusters of size s per lattice site as:

$$n_s = \sum_t g_{st} p^s (1-p)^t \tag{2.6}$$

[5] As they are named in [29]

This formula ia valid for every type of lattice. The difficult part is to calculate g_{st}, i.e. finding all possible configurations and analyze them. That is why, for general s and t, the percolation problem has not yet been solved exactly.

Instead, we have approximate solutions using asymptotic values on the quantities involved. For instance, the perimeter t, averaged over all configurations of a given size s, seems to be proportional to s, for $s \to \infty$. Therefore we can classify these configurations according to the ratio $\alpha = \frac{t}{s}$. For $\alpha < \frac{1-p_c}{p_c}$, the number of lattice animals g_{st} (of size s and perimeter t) varies as $[\frac{(\alpha+1)^{\alpha+1}}{\alpha^{\alpha}}]^s$ for large s. Hence the total number of cluster animals, of size s, $g_s = \sum_t g_{st}$ increases exponentially with s: $g_s \propto s^{-\theta} c^s$ where $c =$ constant. In 2 dimensions $\theta = 1$, in 3 dimensions $\theta = \frac{3}{2}$ and for dimensions > 3 we have $\theta = \frac{5}{2}$. Now from Equation 2.6 we have that the averages over clusters of a fixed size s, correspond (in the limit $p \to 0$) to averages over lattice animals, since the factor $(1-p)^t$ tends to unity and thus can be omitted.

3

Models Of The Small-World

3.1 Milgram's Experiment

One of the first quantitative studies of the structure of social networks was performed in the late 1960's by a Harvard social psychologist named Stanley Milgram [16]. Milgram was interested in the *average distance* between two people and conducted the following experiment:

Milgram distributed letters addressed to a stockbroker acquaintance of his in Boston, Massachusetts, to few hundred randomly selected people in Omaha, Nebraska, considering Boston to be the farthest destination from Nebraska. The letters, targeting the stockbroker in Boston, were to be sent from people of Nebraska to people they knew on a first-name basis. The best strategy was to sent the letter to a person one thought was closer in some social sense (maybe a stockbroker in Boston or a friend in Massachusetts) to the stockbroker in Boston. Meanwhile Milgram was receiving copies of the letters informing him of the intermediate steps the letters followed. The result was that thirty five percent (35%) of the letters reached their destination and the median number of steps these letters took was 5.5 rounding up to 6. A large fraction of the letters never reached their destination and were discarded from the computation of the average distance, so the ones reached their destination only provide an upper bound on the distance.

Though it was implicit in his work, Milgram did not use the term "six degrees of separation". This term was introduced by John Guare in his play titled "Six Degrees of Separation". A character in the play claims that:

> ...Everybody on the planet is separated only by six other people. Six degrees of separation. Between us and everybody else on the planet. The president of the United States. A gondolier in Venice...It's not just the big names. It's anyone. A native in a rain forest. A Tierra del Fuegan. An Eskimo. I am bound to everyone on this planet by a trail of

six people. It's a profound thought.

Milgram generalized the results of his experiment to connect with a chain of six any two randomly chosen people from anywhere in the world. This result is referred to as the *small-world phenomenon*. In a second study, Milgram [14] used essentially the same method to examine the distance of whites in Los Angeles and a mixed white-black target population in New York, and found similar statistics.

Later in 1997, Tjaden and Wasson studied the least distance in the actors graph (The Oracle Of Bacon - http://oracleofbacon.org/). The actors graph linked with an edge any two actors (actresses) appearing in the same movie. The objective was to find the shortest paths between any two actors in the graph. This could be done efficiently by using Kevin Bacon as an intermediate step. This strategy lead to the concept of a "*Bacon number*", meaning the number of links of the shortest path connecting any actor to Kevin Bacon. The distribution of Bacon numbers given in the following table shows that most actors have a small Bacon number:

Bacon number	0	1	2	3	4	5	6	7	8
Number of actors	1	1673	130.852	349.031	84.615	6.718	788	107	11

Table 3.1: Bacon number distribution

The mean distance from Kevin Bacon (as computed using the values from the above table) is 2,94, thus any two actors can be linked by a path through Kevin Bacon in an average of 6 steps.

Albert Barabási and his collaborators, studying the same problem, computed the average distance from each person to all of the others in the actors graph and they found that Rob Steiger, with an average distance of 2,53, was the best choice for an intermediate while Kevin Bacon was found in the 876th place...

Another example of a small-world network is the collaboration graph of mathematics, in which two people are connected if they co-authored a paper. This graph, constructed by Jerrold Grossman [11] in 1997 has 337.454 vertices (84.115 of them isolated) and 496.489 edges. Discarding the isolated vertices the remaining graph has a giant component with 208.200 vertices and 16.883 components with 45.139 vertices. The best intermediate here is Paul Erdös, who wrote more than 500 papers with more than 500 co-authors. The *average* Erdös number is 4.7 while the *largest* Erdös number is 15. Based on a random sample, the average distance between two authors was estimated at 7,37. (These numbers are most likely to change, because in the 1940's 91% of papers in mathematics had only one author, while in the 1990's only 54% did.)

Similar studies that were conducted by Tom Remes in 1997 for baseball players who have played on the same team and by the New York Times (Kirby and Sahre, 1998) with the names of those who had tangled with Monica Lewinsky, also confirmed the surprising result of six degrees of separation.

Besides social networks, small-world properties have also been shown for other networks [32] such as the neural network of the worm Caenorhabditis Elegans (or abbreviated as C.Elegans), where an edge joins two neurons if they are connected by a synapse or a gap junction, the neural network of

the cerebral cortex, as well as the power grid of the western United States, where vertices represent generators, transformers and substations and edges represent high voltage transmission lines between them. Another example of network exhibiting the small-world properties is the World Wide Web studied by Barabási and Albert [3] and Barabási, Albert and Jeong [15] whose vertices are documents and whose edges are links. They estimated that the average distance between vertices scaled with the size of the graph as $0.35 + 2.06 \log n$, thus for $n = 8 \times 10^8$ web pages they obtained 18.59, meaning that any two randomly chosen web pages are on average 19 clicks away from each other.

So far we presented examples of networks showing the small-world properties and reviewed some of their interesting statistics but the small-world phenomenon has not yet been defined precisely. In other words we don't have a specific set of rules a network must obey in order to exhibit the small-world behavior. A first observation is that small-world networks have similar properties with random graphs and to better understand the models of the small-world, some elements of graph theory [8] and random graphs [7] will be inserted here.

3.2 Graph Theoretic Preliminaries

3.2.1 Basic Definitions

Definition. A *graph* is a pair $G = (V, E)$ of sets satisfying $E \subseteq [V]^2, V \cap E = \emptyset$. The elements of E (edges or bonds) are 2-element subsets of V (vertices or nodes or sites).

The number of vertices if a graph G is it's *order* and is denoted by $|G|$. Graphs are finite or infinite according to their order. The number of edges of a graph G is called the *size* of the graph and is denoted by $||G||$.

Definition. A *multigraph* is a pair (V, E) of disjoint sets (of vertices and edges) together with a mapping $E \to V \cup [V]^2$ assigning to every other edge either one or two vertices, its ends. Thus multigraphs can have loops and multiple edges.

Let $G = (V, E)$ be a non-empty graph. The *set of neighbors* of a vertex v in G is denoted by $\Gamma_G(v)$. More generally, for $U \subseteq V$, the neighbors in $V \setminus U$ of vertices in U are called *neighbors* of U and are denoted by $\Gamma(U)$.

Definition. The *degree* $g_G(v) = d(v)$ of a vertex v is the number $|E(v)|$ of edges at v (not for multigraphs) and this equals the number of neighbors of v. A vertex of degree 0 is *isolated*.

Definition. *Minimum-Maximum degree, regular graph, complete graph.*

- The number $\delta(G) = \min\{d(v) \mid v \in V\}$ is the *minimum* degree of the graph G.
- The number $\Delta(G) = \max\{d(v) \mid v \in V\}$ is the *maximum* degree of the graph G.
- If all vertices of G have the same degree k, then G is *k-regular* or just *regular*.
- If all the vertices of G are pairwise adjacent, then G is *complete*.
- A complete graph on n vertices is a K_n and has exactly $\binom{n}{2} = \frac{n(n-1)}{2}$ edges.

- The number $d(G) := \frac{1}{|V|}\sum_{v\in V} d(V)$ is the *average degree* of V.
- Clearly $\delta(G) \leq d(G) \leq \Delta(G)$.

Definition. The *coordination number* z of a vertex v is the number of edges that have v as an endpoint. The coordination number of a vertex v differs from it's degree only in the case of multigraphs, otherwise it is the same.

Definition. The *degree sequence* of an undirected graph is the non-increasing sequence of its vertex degrees. The degree sequence is graph invariant so isomorphic graphs have the same degree sequence.

3.2.2 Characteristic Path Length

A *path* is a non-empty graph $P(V, E)$ of the form $V = \{x_0, x_1, \ldots, x_k\}$, $E = \{x_0x_1, x_1x_2, \ldots x_{k-1}x_k\}$ where the x_i's are all distinct. The vertices x_0 and x_k that are linked by P are called its *ends* and the vertices $x_1, \ldots, x_{k-1}$ are called the *inner* vertices of P.

Definition. The *length of a path* is the number of its edges. A path of length k is denoted by P^k.

If $P = x_0 \ldots x_{k-1}$ is a path and $k \geq 3$ then the graph $R := P + x_{k-1}x_0$ is called a *cycle* or a *ring*. The *length* of a ring its number of edges (or vertices). The ring of length k is denoted by R_k.

Definition. A non-empty graph is called *connected* if any two of its vertices are linked by a path in G.

Definition. The distance $d_G(v, u)$ between two vertices v and u is the length of the shortest $v - u$ path in G. If no such path exists, we set $d_G(v, u) := \infty$.

Definition. The diameter of a graph G is the maximum distance between any two vertices in G and is denoted by $diam(G)$.

As we've seen so far, researchers of small-world networks were mostly interested in the *average* distance in a graph rather than the *maximum* distance, i.e. the diameter. The computation of a closed form expression for the average distance is restricted to connected graphs because of the obvious problems imposed by the infinite path lengths in disconnected graphs.

Definition. The *characteristic path length* of a graph G, denoted by $L(G)$ or just L, is the average distance between any two vertices of G.

The above definition implies that one has to calculate first the shortest path lengths for each vertex $v \in V$ to every other vertex in the graph. That is to calculate $d(v, u), \forall v, u \in V(G), v \neq u$ and then find $\bar{d_v}$ for every $v \in V(G)$. Finally, the characteristic path length is the *median* of all $\{\bar{d_v}\}$. As mentioned before, for various classes of graphs it is difficult to find a closed form expression for the characteristic path length so one has to resort to the explication of upper and lower bounds upon this quantity.

3.2.3 Characteristic path length L of regular lattices

Consider a regular lattice with n vertices where each vertex is connected to all of it's neighbors at distance at most k. That is to say we have a $2k$-regular lattice G of n vertices. It suffices to find the average $d(v,u), \forall v,u \in V(G), v \neq u$ for some vertex $v \in V(G)$, since $\bar{d}_v$ is the same for all vertices in $V(G)$. Starting from vertex v, there are $2k$ vertices for which there exist shortest paths of length 1 to v, $2k$ vertices for which there exist shortest path of length 2 to v, and so on, until we visit all $n-1$ vertices of the graph. These exist unique integers q and r such that $n-1 = 2k \cdot q + r$ with $0 \leq r < 2k$ and $q = \lfloor \frac{n-1}{2k} \rfloor$, meaning that there are $2k$ vertices at distance $\lfloor \frac{n-1}{2k} \rfloor$ from v and the remaining (if any) $r = \text{rem}(n-1, 2k)$ vertices will be at distance $\lfloor \frac{n-1}{2k} \rfloor + 1$ from vertex v. Therefore, the characteristic path length of a $2k$-regular graph, as a function of n, k, is:

$$L(n,k) := \frac{2k \cdot \sum_{i=1}^{\lfloor \frac{n-1}{2k} \rfloor} i + r \cdot (\lfloor \frac{n-1}{2k} \rfloor + 1)}{n-1}, \quad 0 \leq r < 2k \tag{3.1}$$

Example. Consider a regular lattice with $n = 24$ vertices and $k = 3$ as in Figure 3.1, and a vertex $v \in V(G)$. There are $2k = 6$ vertices at distance 1 from v, the next 6 vertices are at distance 2 from v and the next 6 vertices are at distance $\lfloor \frac{n-1}{2k} \rfloor = \lfloor \frac{23}{6} \rfloor = 3$. The remaining $r = \text{rem}(n-1, 2k) = \text{rem}(23, 6) = 5$ vertices will be at distance $\lfloor \frac{n-1}{2k} \rfloor + 1 = \lfloor \frac{23}{6} \rfloor + 1 = 4$.

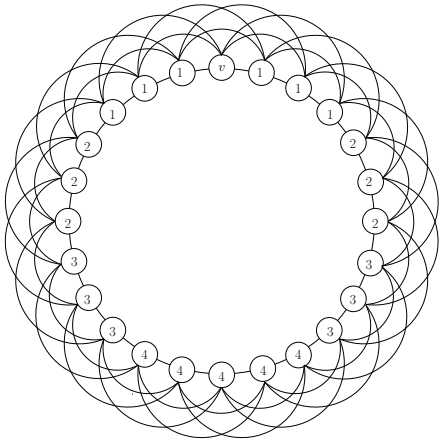


Figure 3.1: A regular lattice with $n = 24$ vertices and $k = 3$

Therefore the characteristic path length in this particular graph is:

$$L(24,3) := \frac{6 \cdot \sum_{i=1}^{\lfloor \frac{23}{6} \rfloor} i + 5 \cdot (\lfloor \frac{23}{6} \rfloor + 1)}{23} = \frac{6 \cdot \sum_{i=1}^{3} i + 5 \cdot (3+1)}{23} = \frac{56}{23} \approx 2,4348$$

In Figure 3.2 we can see the plot of Equation 3.1 (generated with MATLAB©) for lattices with $n = 1 \ldots 100$ vertices and different values of k, varying from 1 to 20. It's easy to see that the characteristic path length grows linearly with the number of vertices and drops as k takes higher values.

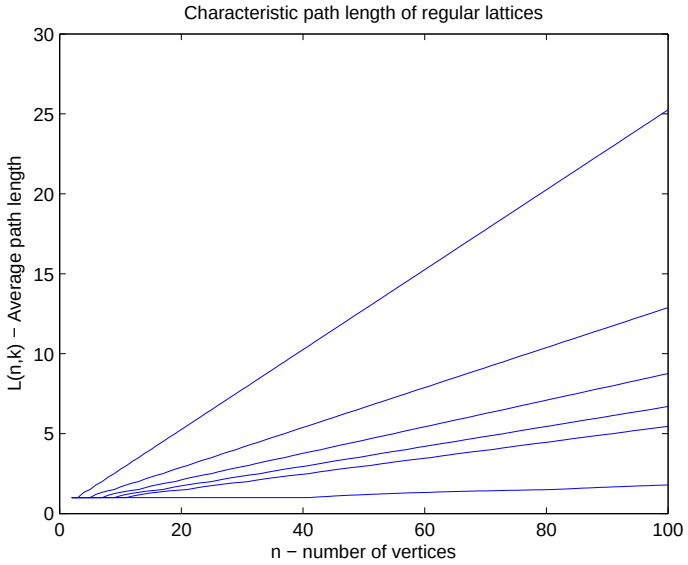


Figure 3.2: Characteristic path length of regular lattices ($n = 1 \dots 100, k = 1, 2, 3, 4, 5, 20$)

3.2.4 Clustering Coefficient

Another essential property of small-world networks is its *clustering*. In social networks, clustering is interpreted as the tendency of one person's circle of acquaintances to overlap. A person's friends are most likely being friends with each other. As a result, social networks, and therefore small-world networks, present some level of *cliquishness* which can be measured by a quantity defined as the *clustering coefficient* [31]. The concept of clustering coefficient has its roots in sociology, appearing under the name "*fraction of transitive triples*" [30].

Definition. The *clustering coefficient* C_v of the *neighborhood* $\Gamma_G(v)$ quantifies the extend to which vertices adjacent to any vertex v are adjacent to each other. More precisely:

$$C_v = \frac{|E(\Gamma_G(v))|}{\binom{|\Gamma_G(v)|}{2}} \tag{3.2}$$

where $|E(\Gamma_G(v))|$ is the number of edges in the neighborhood of v, and $\binom{|\Gamma_G(v)|}{2}$ is the total number of *possible* edges in $\Gamma_G(v)$.

Given $|\Gamma_G(v)|$ vertices, there can be at most $\binom{|\Gamma_G(v)|}{2}$ edges between them. Hence C_v is the fraction of the edges that actually occur in the neighborhood of v divided by the number of all edges that could possibly exist, i.e. as in the complete $K_{|\Gamma_G(v)|}$ subgraph. Equivalently, C_v is the *probability* that two vertices in $\Gamma_G(v)$ will be connected by a path.

Example. Suppose we want to compute the clustering coefficient of vertex v in Figure 3.3.
We have $\Gamma_G(v) = \{v_1, v_2, v_3, v_4\}$, $|\Gamma_G(v)| = 4$ and $|E(\Gamma_G(v))| = 3$ as we can see from Figure 3.3. Therefore the clustering coefficient of vertex v is:

$$C_v = \frac{|E(\Gamma_G(v))|}{\binom{|\Gamma_G(v)|}{2}} = \frac{3}{\binom{4}{2}} = \frac{1}{2}.$$

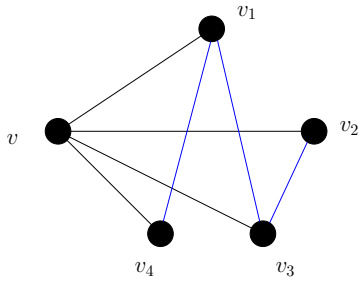


Figure 3.3: Clustering coefficient computation in a graph G

Clustering coefficient - example graphs

By its definition, a clustering coefficient takes values between 0 and 1.

- $C_v = 0$ for a vertex v implies that the neighbors of this vertex have no edges between them. This is expected when v is the center of an asterisk as in Figure 3.4(a).
- $C_v = 1$ for a vertex v implies that every neighbor of v is connected to every other neighbor of v thus forming the complete subgraph as in Figure 3.4(b).

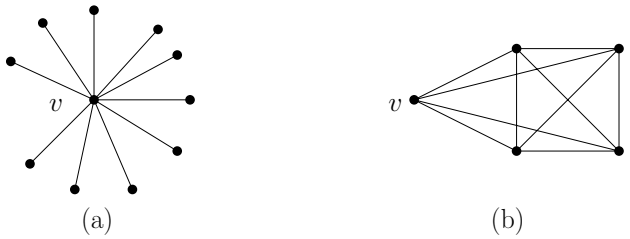


Figure 3.4: Examples of (a) $C_v = 0$ and (b) $C_v = 1$

Definition. The *clustering coefficient* of a *graph* G is $C = \bar{C}_v$ averaged over all vertices $v \in V(G)$.

- $C = 1$ would imply that the corresponding graph consists of $\frac{n}{(k+1)}$ disconnected, but individually complete subgraphs (Figure 3.5(a)) or that the whole graph is a clique (Figure 3.5(b)).
- $C = 0$ would imply that *no* neighbor of any vertex v is adjacent to any other neighbor of v, thus we expect to see a tree-like structure as in Figure 3.6 and / or isolated vertices.

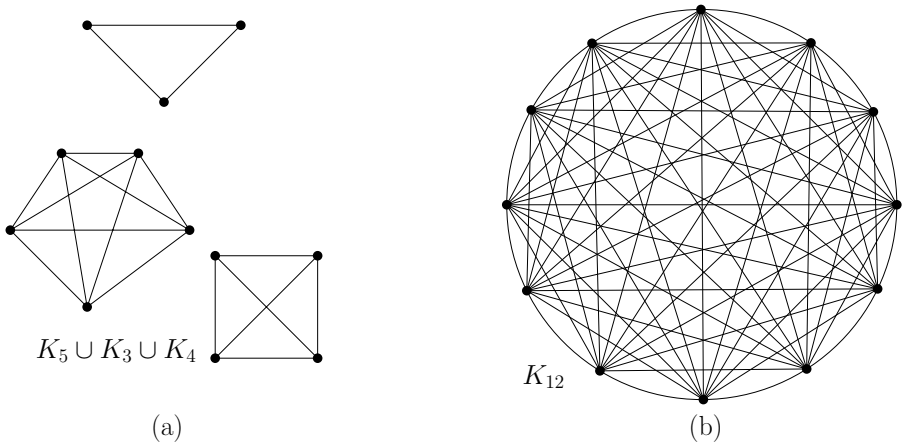


Figure 3.5: Graphs with clustering coefficient $C = 1$

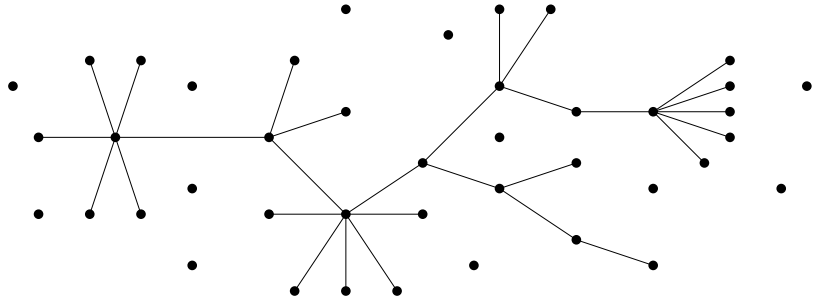


Figure 3.6: A graph with clustering coefficient $C = 0$

3.2.5 Clustering coefficient C in regular lattices

For regular lattices we have an exact calculation of the clustering coefficient C which is a function of the degree (coordination number) z of the vertices. Consider a one-dimensional lattice of infinite length.

- For $z = 2, (k = 1)$ (Figure 3.7) we have $C = 0$. It is obvious that the neighbors of each vertex are not connected with each other.

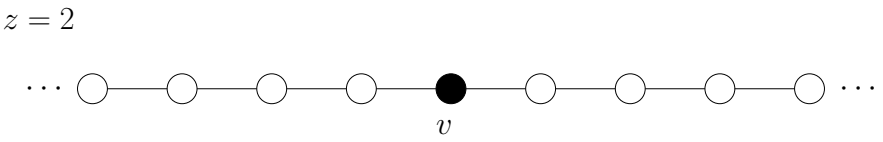


Figure 3.7: A one-dimensional lattice with each site connected to its 2 nearest neighbors

- For $z = 4, (k = 2)$ (Figure 3.8) we have:

$$C_v = \frac{|E(\Gamma_G(v))|}{\binom{|\Gamma_G(v)|}{2}} = \frac{(z-2)+(z-3)}{\binom{z}{2}}$$

- For $z = 6$, $(k = 3)$ (Figure 3.9), we have:

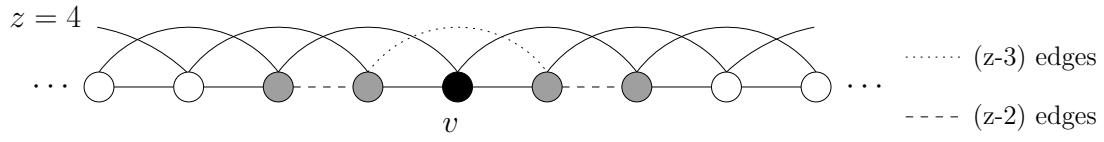


Figure 3.8: A one-dimensional lattice with each site connected to its 4 nearest neighbors

$$C_v = \frac{(z-2)+(z-3)+(z-4)}{\binom{z}{2}}.$$

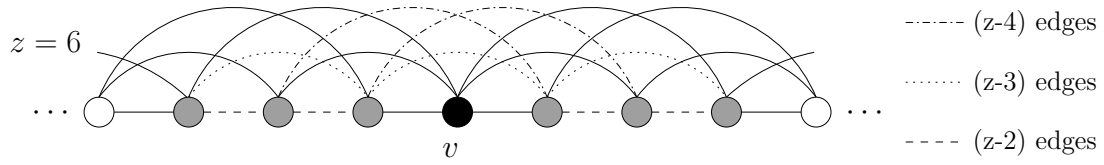


Figure 3.9: A one-dimensional lattice with each site connected to its 6 nearest neighbors

- ... and so on. For general z we have:

$$C_v = \frac{(z-2)+(z-3)+\cdots+[z-(\frac{z}{2}+1)]}{\binom{z}{2}} = \frac{\sum_{i=2}^{\frac{z}{2}+1}(z-i)}{\binom{z}{2}} = \frac{\sum_{i=2}^{\frac{z}{2}+1}(z-i)-(z-1)}{\binom{z}{2}}$$
$$= \frac{\sum_{i=1}^{\frac{z}{2}+1} z - \sum_{i=1}^{\frac{z}{2}+1} i - (z-1)}{\binom{z}{2}} = \frac{z(\frac{z}{2}+1) - \frac{(\frac{z}{2}+1)(\frac{z}{2}+1+1)}{2} - (z-1)}{\frac{z(z-1)}{2}} = \ldots = \frac{3}{4}\frac{(z-2)}{(z-1)}$$

  Averaged over all nodes of the lattice we have $C = \frac{3}{4}\frac{(z-2)}{(z-1)}$. As $z \to \infty$, the clustering coefficient C tends to $\frac{3}{4}$.

3.2.6 Degree Distribution

One additional property small-world networks have, is related to the distribution of the degrees of the network.

Definition. The *degree distribution* $P(k)$ of a network is defined as the fraction of nodes in the network with degree k. Thus if there are n nodes in total in a network and n_k of them have degree k, we have $P(k) = \frac{n_k}{n}$.

For example, in a random graph model where each edge is present with probability p (and absent with probability $1-p$) the probability that a certain node has degree k follows the binomial distribution:

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

This probability represents the number of ways in which k edges can be drawn from a certain node: the probability of k edges is p^k, the probability of the absence of additional edges is $(1-p)^{n-1-k}$ and there are $\binom{n-1}{k}$ ways of selecting the k end points for these edges. To find the degree distribution of the network we need to study the number of nodes X_k with degree k. We will focus on the

probability that X_k takes a certain value, i.e. $P(X_k = r)$. The expected number of nodes with degree k is $E(X_k) = nP(k_i = k) = \lambda_k$, thus the distribution of the X_k values approaches a Poisson distribution with mean value λ_k:

$$P(X_k = r) = e^{-\lambda_k} \frac{\lambda_k^r}{r!}$$

We could say that X_k does not diverge much from the approximative result $X_k = nP(k_i = k)$ which is valid only if nodes are independent. Thus with a good approximation we can say that the degree distribution of a random graph is a binomial distribution $P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$, which for large n can be replaced by a Poisson distribution $P(k) \approx e^{-pn} \frac{(pn)^k}{k!}$. Both binomial and Poisson distributions are strongly peaked about the mean pn and have a large k-tail that decays rapidly as $\frac{1}{k!}$.

However real world networks are mostly found to be very unlike the random graph in their degree distribution. The degrees of the vertices of most networks are highly right skewed which means that their distribution has a long tail of values that are far above the mean [17], [24]. Since the direct histograms are rather noisy, there are two ways to construct a plot of the degree distribution: One way is to construct a histogram in which the bin sizes increase exponentially with degree. For example, the first few bins might cover degree ranges (1, 2-3, 4-7, 8-15 and so on) and then the number of samples in each bin is divided by the width of the bin to normalize the measurement. This method is used when the histogram is to be plotted with a logarithmic degree scale, so that the widths of the bins will appear even. The other way is to make a plot of the *cumulative distribution function*:

$$P_k = \sum_{k_i=k}^{\infty} p_{k_i} \tag{3.3}$$

which is the probability that the degree is greater than or equal to k. It's been shown that for some real world networks the plot of the cumulative distribution function of the degrees is right skewed indicating that the degree distribution approximately follows a power law

$$P_k \backsim \sum_{k_i=k}^{\infty} k_i^{-\alpha} \backsim k^{-(\alpha-1)} \tag{3.4}$$

for some constant exponent α. Such networks are called *scale-free networks*. We can construct networks with a desired power law degree distribution using the following method. We draw a degree sequence $\{k_i\}$ directly from a distribution and we give each vertex i a number k_i of stubs - ends of edges emerging from a vertex. Then we choose pairs of these stubs uniformly at random and join them together to make complete edges (Figure 3.10). When all stubs are used (we restrict their number to be even), the resulting graph is a random member of the ensemble of graphs with the desired degree sequence.
Examples of scale free networks are the citation networks, the World Wide Web, the internet, metabolic networks, telephone call graphs etc.

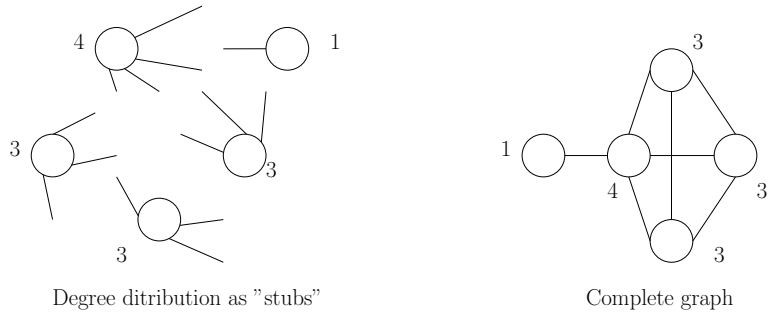


Figure 3.10: Graph construction with a fixed degree sequence

3.2.7 Restrictions

So far we've given definitions and properties of graphs that enable us to represent systems in great detail. However from now on, the class of graphs to which we're interested in, conforms to the following restrictions:

- **Unweighed**: Edges have no weight, meaning that all edges are equivalent and equiprobable as we will see later in random graph models.
- **Simple**: Loops and multiple edges between two vertices are forbidden unless stated otherwise.
- **Sparse**: For an undirected graph, the maximal number of edges is that of the complete graph, i.e. $E(G) = \binom{n}{2} = \frac{n(n-1)}{2}$. The number of edges m in a sparse graph is much less than the number of edges of the corresponding complete graph. Thus $m \ll \frac{n(n-1)}{2}$.
- **Connected**: Any vertex can be reached from any other vertex by traversing a path whose length is finite.

The aforementioned assumptions form a starting point for modelling networks and while they simplify the resulting analysis, they still allow meaningful questions to be asked of a network as a whole. However, the study of small-world networks requires the introduction of new definitions and terminology as well as probabilistic techniques used in the analysis of random graphs. Thus it is important to extend our framework to include definitions and properties from random graph theory that will help us understand the dynamics of small-world graphs.

3.3 Random Graph Preliminaries

Random graph theory was developed in the late 1950s and early 1960s in a series of papers by Erdös and Rènyi [9, 10]. Most of this material is included in Bollobás' standard text [7] published in 1985. It is often helpful to imagine a random graph as a living organism which evolves with time. At first there are n isolated vertices and the edges are added one by one, at each step, according to a random experiment. The nature of this random experiment defines the different classes of random graphs. One objective of random graph theory is to determine at what stage of the evolution, a particular property of a graph is likely to arise.

3.3.1 The Basic Models

Here we will introduce two of the most frequently encountered probability spaces (models) of random graphs. In most cases we consider graphs of n vertices and take $V = \{1, 2, \ldots, n\}$ to be vertex set. The set of all such graphs will be denoted by $\mathcal{G}^n$.

Definition. Consider $V(G) = \{1, 2, \ldots, n\}$ to be the vertex set. $\mathcal{G}(n, M)$ is a random graph model that consists of all graphs with vertex set V having M edges, in which all the graphs have the same probability.

Thus if $N = \binom{n}{2}$, $0 \leq M \leq N$ and $\mathcal{G}(n, M)$ has $\binom{N}{M}$ elements, then every element occurs with probability $\binom{N}{M}^{-1}$.

Definition. The model $\mathcal{G}\{n, P(edge) = p\}$ (or abbreviated as $\mathcal{G}(n, p)$), consists of all graphs with vertex set $V = \{1, 2, \ldots, n\}$ in which every one of the possible $\binom{n}{2}$ edges is chosen independently and with probability $p, 0 \leq p \leq 1$.

In other words if G_0 is a graph in $\mathcal{G}(n, p)$ with vertex set V and m edges, then:

$$P(\{G_0\}) = P(G = G_0) = p^m(1-p)^{N-m}$$

The two models ($\mathcal{G}(n, M)$ and $\mathcal{G}(n, p)$) are practically interchangeable provided that $M \simeq pn$. It is easier to prove theorems in $\mathcal{G}(n, p)$ because the edges are independent whereas in $\mathcal{G}(n, M)$ (where the total number of edges is fixed) there is some dependence of an edge being chosen based on previous choices. This dependance is small, however, and does not affect any important results, so from now on both models will be referred to as *random graphs*.

3.3.2 Properties Of Random Graphs

Random graph theory defines the conditions under which graphs in $\mathcal{G}(n, M)$ and $\mathcal{G}(n, p)$ posses a given property Q, usually in the limit of $n \to \infty$.

Definition. We call a subset Q of $\mathcal{G}$ ($\mathcal{G}(n, M)$ or $\mathcal{G}(n, p)$) a property of graphs of order n. If $G \in Q, H \in \mathcal{G}$ and $G \simeq H$, imply that $H \in G$.

We are mostly interested in the fact that a property Q is a subset of $\mathcal{G}$, thus the statement "G has Q" is equivalent to $G \in Q$. Examples of such properties are: "G is Hamiltonian", i.e. the set $\{G \in \mathcal{G} : G$ is Hamiltonian$\}$ or "G is connected" is the set $\{G \in \mathcal{G} : G$ is connected$\}$.

Definition. A property Q is said to be *monotone increasing* (or simply *monotone*) if whenever $G \in Q$ and $G \subset H$ then also $H \in Q$.

For example the property of a graph containing a certain subgraph, for instance a triangle, is monotone increasing.

Now let Ω_n be a model of random graphs of order n ($\Omega_n = \mathcal{G}(n, M)$ or $\Omega_n = \mathcal{G}(n, p)$). We shall say that "almost every" (a.e.) graph in Ω_n has a certain property Q if $P(Q) \to 1$ as $n \to \infty$. Instead of "almost every" we shall sometimes use "almost all" (a.a.).

A *random graph process* on $V = \{1, 2, \ldots, n\}$ or simply a *graph process* is a Markov chain $\tilde{G} = (G_t)_0^\infty$, whose states are graphs on V. The process starts with an empty graph (isolated vertices) and for $1 \leq t \leq \binom{n}{2}$ the graph G_t is obtained from G_{t-1} by the addition of an edge, all new edges being equiprobable. Then G_t has exactly t edges, thus for $t = \binom{n}{2}$ we have $G_t = K_n$. For $t > \binom{n}{2}$, $G_t = K_n$ as well.

Another different approach to random graph processes is that of being a sequence $(G_t)_t^N = 0$ such that:

- Each G_t is a graph on V.
- G_t has t edges with $t = 0, 1, \ldots, N$, and
- $G_0 \subset G_1 \subset \ldots$

We call G_t the *state* of the process $\tilde{G} = (G_t)_0^N$ at time t. Intuitively, we think of the process $\tilde{G}$ as a living organism which develops by acquiring more and more edges randomly. What we are interested in is to find at what stage of the evolution does a certain property appear. Erdös and Rènyi discovered that most monotone properties appear suddenly: for some function $M = M(n)$ almost no G_M has Q, while for "slightly" larger M almost every G_M has Q.

Definition. Given a monotone increasing property Q, a function $M^*(n)$ is said to be a *threshold function* for Q if:

- $\frac{M(n)}{M^*(n)} \to 0$ implies that almost no G_M has Q, and
- $\frac{M(n)}{M^*(n)} \to \infty$ implies that almost every G_M has Q.

Definition. Suppose Q is a monotone property of graphs. The time at which Q appears is the *hitting time* of Q:

$$\tau = \tau_Q = \tau(\tilde{G}) = \min\{t \geq 0 \mid G_t \text{ has } Q\}.$$

3.3.3 Probability generating functions

Some properties of random graphs can be described with the use of *generating functions* [33]. A *probability generating function* is an alternative representation of a probability distribution. For example, let p_k be the distribution of vertex degrees in a graph. The corresponding generating function is:

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k.$$

Derivatives. This function encapsulates all the information in the original distribution p_k, since we can recover p_k from $G_0(x)$ by simple differentiation:

$$p_k = \frac{1}{k!} \frac{\mathrm{d}^k G_0}{\mathrm{d}x^k}\bigg|_{x=0}$$

One useful property of generating functions is that if the distribution they generate is properly normalized, then:

$$G_0(1) = \sum_k p_k = 1.$$

Moments. With generating functions we can easily calculate the mean of the distribution directly by differentiation:

$$G_0'(1) = \sum_k k p_k = \langle k \rangle.$$

In general, we can calculate any moment of the distribution by taking a suitable derivative:

$$\langle k^n \rangle = \sum_k k^n p_k = \left[\left(x \frac{\mathrm{d}}{\mathrm{d}x} \right)^n G_0(x) \right]_{x=1}.$$

Powers. The last and most important property is that if a generating function generates the probability distribution of some property k of an object, then the sum of that property over m independent such realizations of that object is generated by the mth power of the generating function.

Example. Suppose we choose m vertices at random from a large graph. Then the distribution of the of the sum of the degrees of those vertices is given by the mth power of the generating function, i.e. $[G_0(x)]^m$. To see this we expand the square of the the generating function:

$$\begin{aligned}
[G_0(x)]^2 =& \left[\sum_k p_k x^k \right]^2 \\
=& \sum_{jk} p_j p_k x^{j+k} \\
=& p_0 p_0 x^0 + (p_0 p_1 + p_1 p_0) x^1 + (p_0 p_2 + p_1 p_1 + p_2 p_0) x^2 + (p_0 p_3 + p_1 p_2 + p_2 p_1 + p_3 p_0) x^3 + \ldots
\end{aligned}$$

The coefficients of the powers of x^m in this expression are the sum of all products $p_j p_k$ such that $j + k = m$, hence the probability that the sum of the degrees of the two vertices will be m. This property extends to higher powers of the generating function.

3.4 Models Of The Small-World

What is the connection between random graphs and small-world networks? The most essential property of small-world networks is the small characteristic path length (average distance between any two nodes in the network). Random graphs have also small characteristic path lengths [21]. If a person A on a random graph has z neighbors and each of A's neighbors has also z neighbors, then A has z^2 second neighbors. Extending this argument, A has z^3 third neighbors, z^4 fourth neighbors and so on. Assuming that a person has between 10^2 and 10^3 acquaintances, z^4 is on the order of 10^8 to 10^{12} which is approximately the population of the world. The diameter D can be computed as $z^D = N$ or $D = \frac{\log_N}{\log z}$. The logarithmic increase in the diameter D is typical of the small-world effect. Moreover as N increases, $\log N$ increases only logarithmically, which means that even for very large N the diameter will remain a small number. Another essential property of real world networks (such as the world wide web and the internet) is that they appear to have power law degree distribution and not binomial or Poisson as is the case with Erdös-Rényi random graphs. This means that a small but non-negligible fraction of the vertices in these networks has a very large degree, which has a great effect in the behavior of these networks.

Based on experimental data, real world networks appear to have small-world properties. Our goal is to construct such networks. This construction cannot be based entirely on random graphs. Even if random graphs have small characteristic path lengths (also increasing at *most* logarithmically with the number of nodes of random graphs), they do not show clustering. Two friends of a person A are most likely being friends with each other, while in random graphs this probability is the same as the probability of two *randomly chosen* people being friends with each other. For a random graph the clustering coefficient equals $C = \frac{z}{N}$ which is very small for large networks. Watts & Strogatz [32] calculated the values of the clustering coefficient C_{actual} and the characteristic path length L_{actual} for three different networks: the graph of film actors, the neural network of the worm C.Elegans and the Western Power Grid of the United States. They also calculated the values C_{random} and L_{random} of the corresponding random graphs with the same number of vertices as the aforementioned networks.

Network	N	z	L_{actual}	L_{random}	C_{actual}	C_{random}
Actors Graph	225.226	61	3,65	2,99	0,79	0,00027
Power Grid	4.941	2,67	18,7	12,4	0,080	0,005
C.Elegans	282	14	2,62	2,25	0,28	0,05

Table 3.2: Empirical examples of small-world networks

All three networks listed on Table 3.2, exhibit the small-world phenomenon:

$$L_{actual} \gtrsim L_{random} \quad \text{but} \quad C_{actual} \gg C_{random}.$$

These results suggest that random graphs do not match well the properties of real world networks so it is necessary to find a way of generating graphs that have both properties: small characteristic path length and clustering.

3.4.1 The Watts & Strogatz Small-World Rewiring Model

The first model, proposed by Watts & Strogatz in 1998 [32], interpolates between regularity (lattice) and disorder (random graph). The procedure for generating such graphs is the following:

- We start with a one dimensional lattice, where each site is connected to all sites at distance at most k. Thus each site has $2k$ neighbors, i.e. the initial average coordination number of the graph is $z = 2k$. This construction (Figure 3.11) shows the clustering property: for $k \geq 2$ the neighbors of one site are also neighbors of one another.

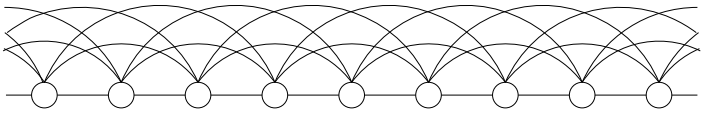


Figure 3.11: A one-dimensional lattice with each site connected to its 6 nearest neighbors

- We apply periodic boundary conditions to the lattice, so that it wraps around on itself (Figure 3.12) in a ring of n sites (R_n).

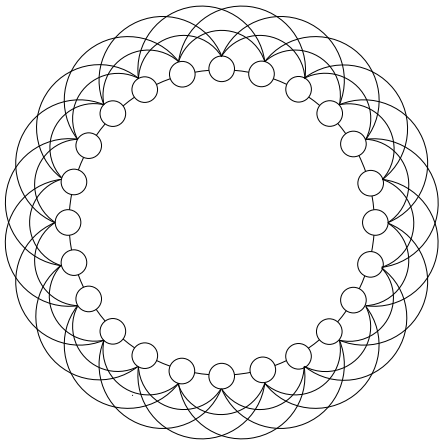


Figure 3.12: A ring lattice with $n = 24$ sites and $z = 6$

- Finally, we choose a site and an edge that connects it to its nearest neighbor in a clockwise sense. With probability ϕ, we reconnect this edge to a site chosen *uniformly at random* over the entire ring, with duplicate edges forbidden. We repeat this process until every site of the ring is considered once. Next we consider the edges that connect sites to their second-nearest neighbor clockwise. We rewire each of these edges with probability p and continue this process until every site is considered once, and so on until every edge is considered once. Since the ring lattice has $\frac{nz}{2}$ edges, the rewiring process will terminate after $\frac{z}{2}$ steps. (Figure 3.13)

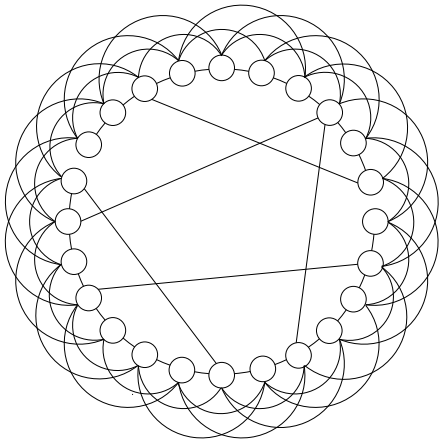


Figure 3.13: The Watts-Strogatz model after rewiring a small fraction of links

Watts & Strogatz showed that for intermediate values of ϕ, the resulting graph is a small-world network. First of all we consider sparse graphs with the least possible number of edges thus we require $n \gg z \ln z \gg 1$, where the condition $z \gg \ln n$ guarantees that a random graph will be connected.

- As $\phi \to 0$ we have $C \sim \frac{3}{4}$ and $L \sim \frac{n}{2z} \gg 1$, which means that we have a highly clustered graph but with very large characteristic path length that grows linearly with n.
- As $\phi \to 1$, $C \approx C_{random} \sim \frac{z}{n} \ll 1$ and $L \approx L_{random} \sim \frac{\ln n}{\ln z}$, thus we have a poorly clustered small-world where L grows only logarithmically with n.

For some broad interval of p we can achieve the desired properties for a small-world network: $L \approx L_{random}$ and $C \gg C_{random}$. The immediate drop of L is caused by the introduction of a few *long-range contacts* or *shortcuts*. For small p, each shortcut has a highly non-linear effect on L whereas C remains practically unchanged. Watts & Strogatz also showed by numerical simulation that $L \approx L_{random}$. For example a random graph with $n = 1000$ and $z = 10$ has $L_{random} = 3,2$. The corresponding ring lattice R_{1000} has $L = 50$. Applying the rewiring model into this ring with probability $p = \frac{1}{4}$ we have $L = 3,6$ which is only slightly larger than L_{random}. Thus the Watts & Strogatz small-world rewiring model appears to show both properties simultaneously: high clustering and small average vertex-vertex distance.

Though the rewiring model has the desired properties for a small-world network, it also has a number of problems. The first problem is that the distribution of shortcuts in the graph is not completely uniform. With duplicate edges forbidden, the new positions for the rewired edges are not all equiprobable. Thus this non-uniformity of the distribution imposes us to work with the average over different realizations of the randomness, a task difficult to perform.

The second problem is that during the *rewiring* process the graph might become disconnected. In that case, distances between the vertices that belong to the disconnected components of the graph are infinite and as a result the characteristic path length of the whole graph becomes infinite. For numerical studies this doesn't appear to be a problem but for analytical work a number of quantities

and expressions are poorly defined. These issues are resolved by a slight modification of the rewiring model.

3.4.2 The Newman-Watts Small-World Model

M. E. J. Newman & D. J. Watts [27] proposed an alternative model: instead of *rewiring* edges we simply *add* shortcuts between pairs of vertices chosen uniformly at random from the ring lattice. In this model duplicate edges and edges that connect a vertex to itself are allowed and no edges are removed from the regular lattice (Example in Figure 3.14).

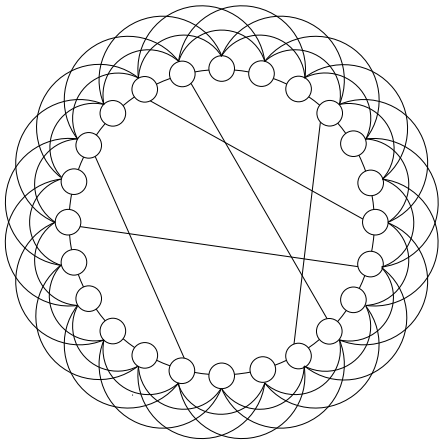


Figure 3.14: A small-world graph with 5 shortcuts added ($n = 24$ and $k = 3$)

For each vertex from the regular lattice we add with probability ϕ one shortcut so that there are ϕn shortcuts on average. The initial average coordination number for the regular lattice is $z = 2k$. Adding shortcuts in the lattice means that a vertex has more endpoints of edges, so the new value of the coordination number becomes:

$$z = \frac{\text{Total number of edges} \cdot 2}{\text{Total number of vertices}} = \frac{(kn + k\phi n) \cdot 2}{n} = 2k(1 + \phi)$$

This model is easier to analyze because it is not possible to split the graph into disconnected components. It has been proved [6] that the characteristic path length L obeys the scaling form $L = \xi \cdot F(\frac{n}{\xi})$, where $F(x)$ is a universal scaling function of its argument x, and ξ is a characteristic length-scale for the model which is assumed to diverge in the limit for small ϕ. Newman & Watts showed that the variable ξ is given by $\xi = \frac{1}{\phi \cdot z}$ for the one-dimensional model. Though it seems that the characteristic path length depends on three parameters (n, z and ϕ) it is actually determined by a single scalar function of a single scalar variable. For $\xi \gg 1$, where it is safe to ignore the scaling in the size of the underlying lattice, and for small ϕ, i.e. when most of a person's acquaintances are local and only a few are long-range, then if we know the form of the scaling function we can thoroughly analyze the model. Newman et al. [25] have calculated the form of the scaling function $F(x)$ using a mean-field-like approximation method which is exact when $x \not\approx 1$. That is:

$$F(x) = \frac{4}{\sqrt{x^2+4x}} \tanh^{-1} \frac{x}{\sqrt{x^2+4x}},$$

but exact analytical calculations for the characteristic path length have been proven very difficult for this particular model. Another problem is the distribution of path lengths in the small-world model. This distribution can be used to provide a simple model of the spread of a disease in a small-world. Newman et al. used the mean-field approximation method to solve this problem too, so for the small-world model with uniformly added shortcuts we have solutions for both the bond and site percolation problems.

3.4.3 Other Models Of The Small-World

Although most research is concentrated on the two aforementioned models, a number of other small-world models have been proposed. It is interesting to see three of these models and investigate some of their properties.

A small-world model with a few highly connected sites

Kasturirangan [12] argued that the small-world phenomenon arises not because there are a few shortcuts in a regular lattice, but because there are a few sites in the network which have unusually high coordination numbers or which are linked to a widely distributed set of neighbors. In the model he proposed we start again with a regular lattice and we add a number of extra vertices in the middle. The next step is to connect these new vertices to a large number of randomly chosen sites from the ring lattice (Figure 3.12).

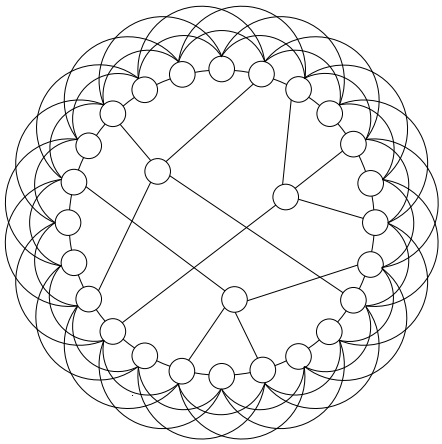


Figure 3.15: A small-world graph with a few highly connected sites

This is an alternative way of introducing shortcuts into a network. This model also shows the small-world effect and has been solved exactly.

Small-world model with a power law degree distribution

Based on the fact that the World Wide Web, which shows the small-world effect, is dominated by a small number of very highly connected sites, Albert et al. [1] proposed a new model similar to the previous one. In this new model the distribution of the coordination number of sites obeys a *power law*, rather than being *bimodal*, as was the case in the previous model. The procedure for generating graphs is as follows:

Starting with a random graph of n sites with average coordination number z, we select a site at random and we add a link between it and another randomly chosen site if that addition would bring the overall distribution of coordination numbers closer to the desired power law; otherwise no link is added. Repeating this procedure, a network is generated with the correct distribution of coordination numbers, yet it remains a random graph. The characteristic path length is small but this type of network doesn't show the clustering property which is essential in small-world graphs.

Kleinberg's small-world model

A third model, proposed by Kleinberg [13], is based on the fact that in social networks people can actually construct short paths given only local information. Such was the case in Milgram's experiment, but in the case of the Watts-Strogatz model no algorithm exists that can find shortest paths given only local information. Kleinberg defined an infinite family of network models that generalize the Watts-Strogatz model, and showed that for one of these models there is a decentralized algorithm capable of finding short paths with high probability. Kleinberg's model is as follows:

We start with a two-dimensional square lattice (Figure 3.16(a)) and we add shortcuts between pairs of vertices i, j with probability which falls off as a power law d_{ij}^{-r} (r-Harmonic distribution) of the distance between them (Figure 3.16(b)). The distance between two vertices i and j with coordinates (x_i, y_i) and (x_j, y_j) respectively is defined as:

$$d((x_i, y_i), (x_j, y_j)) := |x_i - x_j| + |y_i - y_j| \tag{3.5}$$

For arbitrary values of the exponent r it is hard to find a decentralized algorithm that finds shortest paths, but in the case of $r = 2$ such algorithm has been proven to exist.

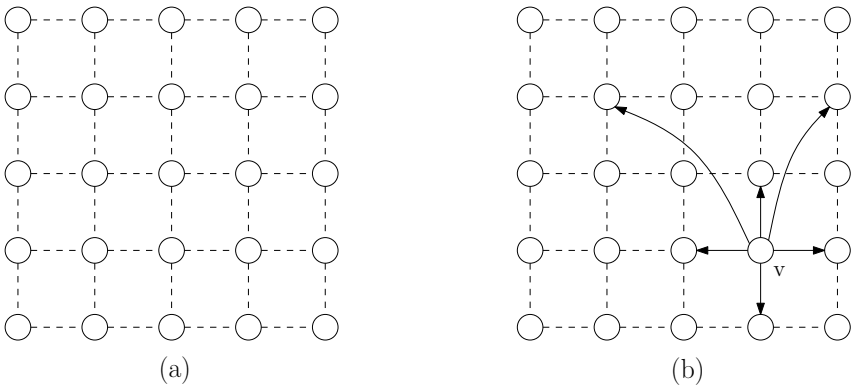


Figure 3.16: Kleinberg's small-world model

One important result regarding this model is that besides the existence of short paths, small worlds are also characterized by the ability to find them without having a global knowledge of the network. From an algorithmic perspective this property should be taken under consideration while constructing a new small-world model.

3.4.4 The r-Harmonic Distribution Model

This last model is in some sense a combination of all the previous models. It is a simplified version of Kleinberg's model, where a ring lattice is used instead of a mesh. A detailed description of the model is as follows.

We start with a *directed* ring of $n+1$ vertices, denoted by R_{n+1}, in which vertices are labelled from 0 to n. The next step is to add *shortcuts* or *long range contacts* between randomly chosen pairs of vertices from the ring lattice.

Consider a graph $G = (V, E)$ and a probabilistic mapping ϕ on the vertices of G such that $\sum_{v \in V} \phi(u, v) = 1$ for all $u \in V$, i.e. each vertex $u \in V$ has an associated probability distribution $\phi(u, \cdot)$. Based on the type of this distribution we may obtain a variety of different models. Motivated by Kleinberg's research we will use the *r-harmonic* $r \geq 0$ distribution. Two examples of Harmonic distributions are:

- For $r = 1$, we have the *uniform* distribution where $\phi(u, v) = \frac{1}{n}$, and
- For $r = 1 - \frac{\log 0.80}{\log 0.20}$, we have the *Zipf* distribution.

Given two vertices u and v the probability for u to have v as long range contact in a graph G is given by:

$$\phi_r(u, v) = \frac{d(u, v)^{-r}}{\sum_{w \neq u} d(u, w)^{-r}} \tag{3.6}$$

where $d(\cdot, \cdot)$ is the distance function of the graph. In the directed labelled ring R_{n+1} the distance between two vertices with labels i and j is defined as $d(i, j) = (j - i) \mod n+1$, thus the probability in Equation 3.6 can be simplified.

Now consider the *r-harmonic random variable* H_r, which takes values in $\{1, 2, \ldots, n\}$. This random variable has probability distribution defined by:

$$\mathbf{Pr}(\{H_r = x\}) = \frac{x^{-r}}{H_n^{(r)}},$$

where $H_n^r = \sum_{i=1}^{n} i^{-r}$ is the r-harmonic number of order n. If the ring R_{n+1} is augmented using the r-harmonic mapping ϕ_r, then given two vertices with labels i and j the probability for i to have j as long range contact in (R_{n+1}, ϕ_r) is given by:

$$\phi_r(i, j) = \frac{((j - i) \bmod n + 1)^{-r}}{H_n^r} \tag{3.7}$$

4

Percolation On Small-World Networks

In the previous section, a number of small-world models were presented along with their most important properties. Those properties however describe the static structure of the networks. To better understand the benefits of each model we must extensively study their dynamics. There are a number of dynamical systems that can be defined on small-world networks such as networks of coupled oscillators or cellular automata. Here we will concentrate on *epidemic* or *disease propagation* models on small-world graphs which are essentially *percolation processes*. In epidemiology there are two parameters of interest: *susceptibility*, i.e. the probability that an individual exposed to a disease will contract it, and *transmissibility*, i.e. the probability that a healthy but susceptible individual will contract the disease once it has a contact with an infected individual.

Infected individuals are represented by occupied sites on a small-world graph and the disease spreads along the bonds which are represented by edges between the sites. A disease begins with a single infected individual. We will study two extremes of this model:

- In the first case only a fraction p of the individuals are susceptible but if an individual gets infected, all of its susceptible neighbors will contract the disease. In percolation terms this is the *site* percolation problem.
- In the second case all individuals are susceptible and there is a probability p that an infected individual will transmit the infection to a neighbor. This corresponds to the *bond* percolation problem.

We will investigate both the site and bond percolation problems first on the Newman-Watts model and next on the r-Harmonic distribution model.

4.1 Percolation On The Newman-Watts Small-World Model

Several methods have been proposed for the study of disease propagation on random networks. In random graph theory percolation happens on a network when a *giant component* appears, i.e. a connected component whose size approaches the size of the whole graph. A disease outbreak which starts with a single individual will spread only within connected components, thus at a certain value of the percolation threshold p_c the system undergoes a phase transition which is the onset of epidemic behavior. Epidemics in random networks were studied with the generating function method [22], [26], [23]. Moore and Newman studied percolation on small world networks using a transfer matrix method [18] and later using a generating functions method [19]. Both methods will be presented in the following text.

4.1.1 Site Percolation

Generating Functions Method

The basic idea of the generating function method is to find the distribution of *local clusters* (defined later in this text) and calculate how these local clusters are joined together with shortcuts to form *connected clusters*. Next we find a closed form expression for the mean connected cluster size. When this cluster size diverges, we are right above the phase transition where a giant connected component forms. This is exactly the point where we can compute the percolation threshold of the system.

Initial structure. We consider a one-dimensional small-world graph with L sites arranged on a regular ring lattice with periodic boundary conditions. Each site is connected to all sites at distance at most k, thus the initial coordination number of each site is $z = 2k$. A number of shortcuts are now added to the graph between pairs of vertices chosen uniformly at random. Let ϕ be the average number of shortcuts per bond on the underlying lattice. Therefore we have a total of $k\phi L$ shortcuts. The probability of two randomly chosen sites having a shortcut between them is:

$$\begin{aligned}\psi &= \Pr[\text{Two randomly chosen sites have a shortcut between them}] \\ &= 1 - \Pr[\text{A specific pair of sites have no shortcut between them}]\end{aligned}$$

In the Newman-Watts small-world model, when adding shortcuts, loops and duplicates are allowed, so there are L^2 "pairs" of sites which might have a shortcut "between" them. Consider a specific pair (l_i, l_j) of sites we don't want to connect with a shortcut. Since the graph is undirected we also don't want to have a shortcut between the pair (l_j, l_i) of sites. Therefore we have $(L^2 - 2)$ "unavailable" positions of shortcuts out of the overall L^2. Finally we shall distribute the $k\phi L$ shortcuts to the available positions with $(L^2 - 2)^{k\phi L}$ ways. Hence the probability ψ is:

$$\psi = 1 - [\frac{L^2 - 2}{L^2}]^{k\phi L} \approx \frac{2k\phi}{L}, \text{ for large } L. \tag{4.1}$$

We can assume that every shortcut leads to a different local cluster for large L since the probability of two shortcuts connecting the same pair of local clusters falls off as L^{-1}.

The next step is to show the susceptible individuals. As mentioned earlier, the contact between an infected and a healthy but susceptible individual results in the latter contracting the disease. Less than 100% of the individuals are susceptible, therefore we represent them with a fraction p of occupied (colored) sites on the graph. An example of this structure is presented in Figure 4.1.

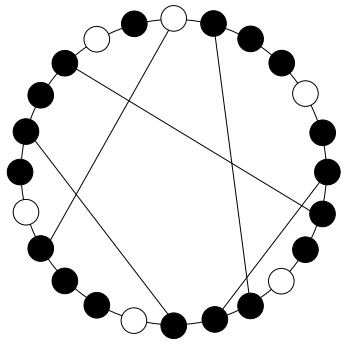


Figure 4.1: A small-world with $L = 24$ sites, 4 shortcuts and $p = \frac{3}{4}$ susceptible individuals

The occupied sites form a number of clusters. First the occupied sites which are connected with the nearest neighbor bonds on the underlying one-dimensional lattice form *local clusters*. These local clusters are connected together by shortcuts to form the *connected clusters* of the small-world network. Thus a connected cluster (circle) is equal to a single local cluster (square) with any number of connected clusters attached to it by a single shortcut. This recursive tree-like structure is shown in Figure 4.2.

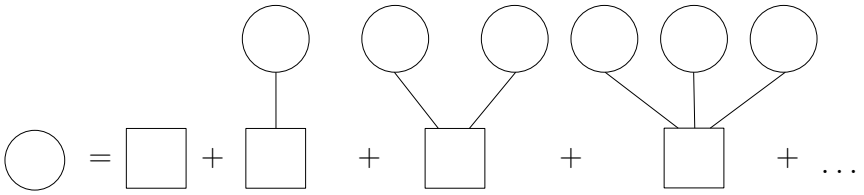


Figure 4.2: Graphical representation of a cluster of connected sites.

Local Clusters. First we have to calculate the number of local clusters of length n. This is also the probability $P_0(n)$ that a randomly chosen site belongs to a local cluster of size n. We define the following procedure:

- Start with an occupied cluster. This is the initial local cluster of size 1.
- On the ring lattice, check the neighbors of this site at distance at most k. If these neighbors are occupied (with probability p) then add their number (how many they are) to the initial cluster.
- Repeat the previous step until all of the neighbors of the sites of the local cluster formed so far are unoccupied (with probability $(1 - p)$).

We may think of this process as following a geometric distribution[1] which starts with an occupied site with probability p and terminates with success probability $(1-p)^k$ after $(n-1)$ steps.

- For $n = 0$ the average number of local clusters of length n is:

$$P_0(n) = 1 - p$$

- For general k and $n > 0$ we have:

$$\begin{aligned} P_0(n) &= (1-p)^{2k} p (1-(1-p)^k)^{n-1} n \\ &= (1-p)^2 p q^{n-1} n \end{aligned}$$

where $q = 1 - (1-p)^k$.

Here the process starts with an occupied site with probability p, for the next $(n-1)$ steps this process "fails" with probability $1-(1-p)^k$ and terminates with success probability $(1-p)^k$.

Thus we have:

$$P_0(n) = \begin{cases} 1-p & \text{for } n = 0 \\ (1-p)^2 p q^{n-1} n & \text{for } n > 0 \end{cases} \tag{4.2}$$

Let $H_0(z)$ be the generating function for the local clusters. Then:

$$H_0(z) = \sum_{n=0}^{\infty} P_0(z) z^n \tag{4.3}$$

Using 4.2 we can calculate this generating function.

$$\begin{aligned} H_0(z) &= 1 - p + \sum_{n=1}^{\infty} (1-p)^2 p q^{n-1} n z^n \\ &= 1 - p + (1-p)^2 \frac{p}{q} \sum_{n=1}^{\infty} n (qz)^n \qquad \left[\sum_{i=0}^{\infty} i x^i = \frac{x}{(1-x)^2} \right] \\ &= 1 - p + (1-p)^2 \frac{p}{q} \frac{qz}{(1-qz)^2} \end{aligned}$$

Hence:

$$H_0(z) = \sum_{n=0}^{\infty} P_0(z) z^n = 1 - p + pz \frac{(1-p)^2}{(1-qz)^2} \tag{4.4}$$

[1] Consider a sequence of independent Bernoulli trials with success probability p (and failure probability q the same for each trial). Let X be the number of trials till we succeed. Then $P(X = x) = pq^{x-1}$.

Connected Clusters. Now let $P(n)$ be the probability that a randomly chosen site belongs to a *connected* cluster of n sites. This is also the probability that a disease outbreak starting with a randomly chosen individual will affect n people. Since $P(n)$ is difficult to calculate, we use the generating function method. Let $H(z)$ be the generating function for the probability $P(n)$. Then:

$$H(z) = \sum_{n=0}^{\infty} P(n) z^n \tag{4.5}$$

Since the probability of two shortcuts connecting the same pair of local clusters falls off as L^{-1}, this means that each connected cluster consists of a local cluster with $m \geq 0$ shortcuts leading from it to m connected clusters. Thus $H(z)$ satisfies the Dyson-equation-like iterative condition, which we can write self-consistently as:

$$H(z) = \sum_{n=0}^{\infty} P_0(n) z^n \sum_{m=0}^{\infty} P(m|n) [H(z)]^m \tag{4.6}$$

$P(m|n)$ is the conditional probability of there being exactly m shortcuts emerging from a local cluster of size n. Since there are ϕkL shortcuts in the network, there will be $2\phi kL$ ends of shortcuts. Therefore $P(m|n)$ is given by the binomial:

$$P(m|n) = \binom{2\phi kL}{m} \left[\frac{n}{L}\right]^m \left[1 - \frac{n}{L}\right]^{2\phi kL - m} \tag{4.7}$$

We take m ends of shortcuts out of the overall $2\phi kL$, we connect them to a local cluster of size n (with m ways) and we connect the remaining $2\phi kL - m$ ends of shortcuts to the remaining local clusters of sizes other than n (with $2\phi kL - m$ ways).
Using Equation 4.7, Equation 4.6 becomes:

$$\begin{aligned}
H(z) &= \sum_{n=0}^{\infty} P_0(n) z^n \sum_{m=0}^{\infty} P(m|n) [H(z)]^m \\
&= \sum_{n=0}^{\infty} P_0(n) z^n \sum_{m=0}^{\infty} \binom{2\phi kL}{m} \left[\frac{n}{L}\right]^m \left[1 - \frac{n}{L}\right]^{2\phi kL - m} [H(z)]^m \\
&= \sum_{n=0}^{\infty} P_0(n) z^n \sum_{m=0}^{\infty} \binom{2\phi kL}{m} \left[\frac{n}{L-n} H(z)\right]^m \left[\frac{L}{L-n}\right]^{-2\phi kL} \\
&= \sum_{n=0}^{\infty} P_0(n) z^n \left[\frac{L}{L-n}\right]^{-2\phi kL} \sum_{m=0}^{\infty} \binom{2\phi kL}{m} \left[\frac{n}{L-n} H(z)\right]^m, \quad \left[\text{since} \sum_{i=0}^{\infty} \binom{n}{i} x^i = (1+x)^n\right] \\
&= \sum_{n=0}^{\infty} P_0(n) z^n \left[\frac{L}{L-n}\right]^{-2\phi kL} \left[1 + \frac{n}{L-n} H(z)\right]^{2\phi kL}
\end{aligned}$$

Therefore:

$$H(z) = \sum_{n=0}^{\infty} P_0(n) z^n \left[1 + (H(z) - 1)\frac{n}{L}\right]^{2\phi kL} \tag{4.8}$$

For large L we can approximate this expression with:

$$H(z) = \sum_{n=0}^{\infty} P_0(n)[ze^{2k\phi(H(z)-1)}]^n \tag{4.9}$$

It is easy to see that $H(z)$ is equal to $H_0(z)$ if we replace z with $ze^{2k\phi(H(z)-1)}$. Thus:

$$H(z) = H_0(ze^{2k\phi(H(z)-1)}) \tag{4.10}$$

We can calculate $H(z)$ by iterating this equation starting with $H(z) = 1$. The next step is to calculate the mean outbreak size which is given by the first derivative of H at $z = 1$:

$$\begin{aligned}
H'(1) &= \frac{d}{dz}[H_0(ze^{2k\phi(H(z)-1)})]\Big|_{z=1} \\
&= H_0'(1)\frac{d}{dz}[ze^{2k\phi(H(z)-1)}]\Big|_{z=1} \\
&= H_0'(1)\frac{d}{dz}[e^{2k\phi(H(z)-1)} + ze^{2k\phi(H(z)-1)}2k\phi H'(z)]\Big|_{z=1} \\
&= H_0'(1)(1 + 2k\phi H'(1))
\end{aligned}$$

Therefore:

$$H'(1) = \frac{H_0'(1)}{1 - 2k\phi H_0'(1)} \tag{4.11}$$

We can calculate the first derivative of $H_0(z)$ at $z = 1$ from Equation 4.4:

$$\begin{aligned}
H_0'(1) &= \frac{d}{dz}\left[1 - p + pz\frac{(1-p)^2}{(1-qz)^2}\right]\Big|_{z=1} \\
&= p(1-q)^2\frac{d}{dz}\left[\frac{z}{(1-qz)^2}\right]\Big|_{z=1} \\
&= p(1-q)^2\frac{(1+qz)}{(1-qz)^3}\Big|_{z=1} \\
&= p\frac{1+q}{1-q}
\end{aligned} \tag{4.12}$$

Equation 4.11 thus gives:

$$H'(1) = \frac{\frac{p(1+q)}{(1-p)}}{1 - 2k\phi\frac{p(1+q)}{(1-q)}} = \frac{p(1+q)}{1-q-2k\phi p(1+q)} = \frac{p(2-(1-p)^k)}{(1-p)^k - 2k\phi p(2-(1-p)^k)} \tag{4.13}$$

The last step is to calculate the percolation threshold. The mean outbreak size diverges at the percolation threshold $p = p_c$. This happens when the denominator of Equation 4.13 is zero, i.e.:

$$1 - q_c - 2k\phi p_c(1+q_c) = 0 \rightarrow \phi = \frac{1-q_c}{2kp_c(1+q_c)} = \frac{(1-p_c)^k}{2kp_c(2-(1-p_c)^k)} \tag{4.14}$$

- For $k = 1$, the percolation threshold is the solution of a quadratic equation:

$$2\phi p_c^2 + (2\phi + 1)p_c - 1 = 0 \rightarrow p_c = \frac{\sqrt{4\phi^2 + 12\phi + 1} - 2\phi - 1}{4\phi} \tag{4.15}$$

- For general k the percolation threshold is the solution of a polynomial of order $k + 1$.

Transition Matrix Method

The basic idea of this method is to consider cluster growth as a stochastic process evolving it time. Starting with a particular local cluster, we add to it all the other local clusters that can be reached from it by a single shortcut. Then we add all the local clusters that can be reached from the newly added local clusters and so on until a connected cluster is formed and there are no more shortcuts that lead to new local clusters.

This process has the Markov property, i.e. the probability that a process is in a particular state y at time t depends only on its state x at time $t - 1$. This probability is denoted by $P(x, y)$ and is independent of time t. The values $P(x, y)$ are called *transition probabilities*.

$$\mathbf{Pr}[X_t = y \mid X_{t-1} = x, X_{t-2}, \ldots, X_0] = \mathbf{Pr}[X_t = y \mid X_{t-1} = x] = P(x, y)$$

At each step of this process we add new local clusters to the overall connected cluster based on our current position in the lattice. We assume that every shortcut leads us to a different local cluster for large L. We can model this process using a *transition matrix* $\mathbf{M}$ whose elements are the transition probabilities of this process. Let $\mathbf{v}$ be a column vector whose elements v_i are equal to the probability that a local cluster of size i has just been added to the overall connected cluster. We wish to calculate the values of this vector during the evolution of this process in time. Let $\mathbf{v}^{(0)}$ be the initial distribution vector. In the next time step we have:

$$\mathbf{v}^{(1)} = \mathbf{M}\mathbf{v}^{(0)} \quad \text{or} \quad v_i' = \sum_j M_{ij} v_j \tag{4.16}$$

This transition matrix is independent of the time step and is repeatedly applied to the vector of probabilities $\mathbf{v}$ until there are no more local clusters to be added to the overall connected cluster. This happens when we reach the percolation threshold (equilibrium state). In this state, vector $\mathbf{v}'$ is stationary.

The elements of matrix[2] $\mathbf{M}$ are:

$$M_{ij} = N_i(1 - (1 - \psi)^{ij}),$$

where N_i is the average number of local clusters of size i and $1-(1-\psi)^{ij}$ is the probability of having a shortcut from a local cluster of size i to one of size j, since there are ij possible pairs of sites by which these can be connected.

[2]The matrix $\mathbf{M}$ is not stochastic, i.e. its rows do not sum to unity, yet its entries are strictly positive.

From the previous section we know that the probability that a randomly chosen site belongs to a local cluster of size n is:

$$P_0(n) = \begin{cases} 1-p, & \text{for } n = 0 \\ (1-p)^2 pq^{n-1} n, & \text{for } n > 0 \end{cases}$$

Therefore the average number of local clusters of size i is $N_i = \frac{P_0(i)}{i} L$ or:

$$N_i = \begin{cases} (1-p)^2 p^i L, & \text{for } k = 1 \\ (1-p)^2 pq^{i-1} L, & \text{for } k > 1 \end{cases} \tag{4.17}$$

where $q = (1-(1-p)^k)$. As this process evolves in time we have:

$$\begin{aligned} \mathbf{v}^{(1)} &= \mathbf{M}\mathbf{v}^{(0)} \\ \mathbf{v}^{(2)} &= \mathbf{M}\mathbf{v}^{(1)} = \mathbf{M}^2\mathbf{v}^{(0)} \\ &\vdots \\ \mathbf{v}^{(n)} &= \mathbf{M}^n\mathbf{v}^{(0)} \\ &\vdots \end{aligned}$$

At the percolation threshold (after a finite number of steps) this process has reached the equilibrium state where:

$$\mathbf{v}' = \mathbf{M}\mathbf{v}'$$

and no matter how many times we apply the transition matrix $\mathbf{M}$ on vector $\mathbf{v}'$, $\mathbf{v}'$ remains unchanged. It is easy to see that this vector $\mathbf{v}'$ is the right eigenvector of the transition matrix $\mathbf{M}$.

Now consider the largest eigenvalue of $\mathbf{M}$, i.e. the largest value of λ for which $(\mathbf{M} - \lambda\mathbf{I})\mathbf{v}' = \mathbf{0}$.

- If $\lambda < 1$, applying matrix $\mathbf{M}$ to vector $\mathbf{v}'$, makes $\mathbf{v}'$ tend to zero and as a result the rate at which new local clusters are added falls off exponentially and the connected clusters are finite with exponential size distribution.
- If $\lambda > 1$, applying matrix $\mathbf{M}$ to vector $\mathbf{v}'$, makes $\mathbf{v}'$ growing until the size of the overall cluster becomes limited by the size of the whole system.
- Thus percolation threshold occurs at $\lambda = 1$.

For finite L it's difficult to calculate the largest eigenvalue λ of $\mathbf{M}$, but for $L \to \infty$, ϕ being constant and $\psi \to 0$ we can approximate $\mathbf{M}$ by:

$$M_{ij} = ij\psi N_i \tag{4.18}$$

Replacing M_{ij} in Equation 4.16 and setting $v_i' = \lambda v_i$ we have:

$$\lambda v_i = i\psi N_i \sum_j j v_j \tag{4.19}$$

Thus the eigenvectors of $\mathbf{M}$ have the form $v_i = C\lambda^{-1} i\psi N_i$ where $C = \sum_j j v_j$ is a constant. Eliminating C we have:

$$\lambda = \psi \sum_j j^2 N_j \tag{4.20}$$

- For $k = 1$, i.e. $N_i = (1-p)^2 p^i L$ we have:

$$\begin{aligned}
\lambda &= \psi \sum_j j^2 N_j = \psi \sum_j j^2 (1-p)^2 p^j L = \psi (1-p)^2 L \sum_j j^2 p^j \\
&= \psi (1-p)^2 L \left(\sum_j j^2 p^j - j p^j + \sum_j j p^j \right) = \psi (1-p)^2 L \left(p^2 \sum_j \frac{d^2(p^j)}{dj^2} + p \sum_j \frac{d(p^j)}{dj} \right) \\
&= \psi (1-p)^2 L \left(p^2 \frac{d^2(\sum_j p^j)}{dj^2} + p \frac{d(\sum_j p^j)}{dj} \right) = \psi (1-p)^2 L \left(p^2 \frac{2}{(1-p)^3} + p \frac{1}{(1-p)^2} \right) \\
&= \psi L p \frac{1+p}{1-p} = \frac{2k\phi}{L} L p \frac{1+p}{1-p} = 2k\phi p \frac{1+p}{1-p}
\end{aligned}$$

  For $\lambda = 1$ we get the value of p at the percolation threshold, i.e. $p = p_c$.

$$1 = 2k\phi p \frac{1+p}{1-p} \rightarrow p_c = \frac{\sqrt{4\phi^2 + 12\phi + 1} - 2\phi - 1}{4\phi}$$

- For $k > 1$ we have:

$$\lambda = \psi L p \frac{1+q}{1-q} = 2k\phi p \frac{2 - (1-p)^k}{(1-p)^k} \tag{4.21}$$

  At the percolation threshold ($\lambda = 1$):

$$\phi = \frac{(1-p_c)^k}{2kp_c(2 - (1-p_c)^k)}$$

  Thus the percolation threshold is the solution of a polynomial of order $k+1$ in agreement with the result in Equation 4.14.

Remarks From Equation 4.21 we can see that the value of the percolation threshold depends on the size of the ring lattice L and the probability ψ of having a shortcut between two nodes. Thus the same method can be applied in other small world models were the probability of having a shortcut between two nodes follows some distribution.

4.1.2 Bond Percolation

In this section we will study the bond percolation problem on the Newmann-Watts small world network. The bond percolation problem is equivalent to the disease propagation problem where all individuals are susceptible but transmission takes place with less than 100% efficiency. In this model, when a sufficient fraction p_c of the bonds of the network are occupied, they form a giant component whose size scales extensively with the size of the network. The fraction p of occupied bonds is the transmissibility of the disease. We will study two separate cases: for $k = 1$ and for $k > 1$.

Generating Functions Method. For $k = 1$, the probability $P_0(n)$ that a randomly chosen site belongs to a *local* cluster of size n is:

$$P_0(n) = \begin{cases} 0 & \text{for } n = 0 \\ (1-p)^2 p^{n-1} n & \text{for } n > 0 \end{cases} \tag{4.22}$$

where p is the bond occupation probability. Here a local cluster of n sites consists of $n-1$ occupied bonds with two unoccupied bonds at either end. Thus $P(n)$ (the probability that a randomly chosen site belongs to a *connected* cluster of size n) can be computed through the generating function of Equation 4.6 as:

$$H(z) = \sum_{n=0}^{\infty} P_0(n) z^n \sum_{m=0}^{\infty} P(m|n)[H(z)]^m \tag{4.23}$$

Now a shortcut must be an open bond with probability p thus Equation 4.8 is slightly modified (replace ϕ with $p\phi$). Thus:

$$H(z) = \sum_{n=0}^{\infty} P_0(n) z^n \left[1 + (H(z) - 1)\frac{n}{L}\right]^{2p\phi L} = \sum_{n=0}^{\infty} [pze^{2p\phi(H(z)-1)}]^n \tag{4.24}$$

where $H(z)$ is equal to $H_0(z)$ with $z \to ze^{2p\phi(H(z)-1)}$. Therefore:

$$H(z) = H_0(ze^{2p\phi(H(z)-1)}) \tag{4.25}$$

The mean outbreak size is given by the first derivative of H, i.e.:

$$\begin{aligned} H'(1) &= \frac{d}{dz}[H_0(ze^{2p\phi(H(z)-1)})]\Big|_{z=1} \\ &= H_0'(1)\frac{d}{dz}[ze^{2p\phi(H(z)-1)}]\Big|_{z=1} \\ &= H_0'(1)\frac{d}{dz}[e^{2p\phi(H(z)-1)} + ze^{2p\phi(H(z)-1)}2p\phi H'(z)]\Big|_{z=1} \\ &= H_0'(1)(1 + 2p\phi H'(1)) \end{aligned}$$

Solving for $H'(1)$ we have:

$$H'(1) = \frac{H_0'(1)}{1 - 2p\phi H_0'(1)} \tag{4.26}$$

In order to calculate $H_0(z)$ first we have to calculate $H_0(z)$:

$$\begin{aligned} H_0(z) &= 0 + \sum_{n=1}^{\infty} P_0(z) z^n = \sum_{n=1}^{\infty} (1-p)^2 p^{n-1} n z^n \\ &= \frac{(1-p)^2}{p} \sum_{n=1}^{\infty} n (pz)^n \qquad\qquad \left[\sum_{i=0}^{\infty} i x^i = \frac{x}{(1-x)^2} \right] \\ &= \frac{(1-p)^2}{p} \frac{pz}{(1-pz)^2} = z \frac{(1-p)^2}{(1-pz)^2}. \end{aligned}$$

Therefore:

$$\begin{aligned} H_0'(1) &= \frac{d}{dz} \left[z \frac{(1-p)^2}{(1-pz)^2} \right] \Bigg|_{z=1} = (1-p)^2 \frac{d}{dz} \left[\frac{z}{(1-pz)^2} \right] \Bigg|_{z=1} \\ &= (1-p)^2 \left[\frac{(1+pz)}{(1-pz)^3} \right] \Bigg|_{z=1} = (1-p)^2 \left[\frac{(1+p)}{(1-p)^3} \right] \\ &= \frac{1+p}{1-p} \end{aligned} \tag{4.27}$$

Using Equation 4.27, Equation 4.26 gives:

$$H'(1) = \frac{H_0'(1)}{1 - 2p\phi H_0'(1)} = \frac{\frac{1+p}{1-p}}{1 - 2p\phi \frac{1+p}{1-p}} = \frac{(1+p)}{1-p-2p\phi(1+p)} \tag{4.28}$$

The onset of epidemic behavior (where $p = p_c$) occurs at the zero of the denominator of Equation 4.28 or when:

$$\phi = \frac{1-p_c}{2p_c(1+p_c)}$$

Solving for p_c we have the quadratic Equation:

$$2\phi p_c^2 + (2\phi + 1) p_c - 1 = 0$$

with solution:

$$p_c = \frac{\sqrt{4\phi^2 + 12\phi + 1} - 2\phi - 1}{4\phi} \tag{4.29}$$

which is exactly the same solution we got for the site percolation problem on the same network for $k = 1$.

Bond percolation when $k > 1$. For $k > 1$ calculating the average number of local clusters is a difficult so we will solve the case $k = 2$. In the bond percolation problem all individuals (nodes) are susceptible but the bonds between them are open or closed based on some probability p. A single node with closed bonds to it's right and left neighbors is a local cluster of size 1 as shown in Figure 4.3. Moreover every shortcut has meaning only if it is an open bond with probability ϕp.

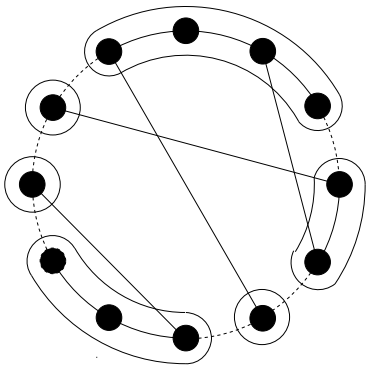


Figure 4.3: Local clusters for the bond percolation problem.

Now let Q_i be the probability that a given site n and it's left neighbor site $n-1$ are part of the *same* local cluster of size i when only bonds to the left of site n are taken into account. Let Q_{ij} be the probability that sites n and $n-1$ are parts of two *separate* local clusters of sizes i and j respectively again when only bonds to the left of n are considered. Let us consider site $n+1$. This can be possibly connected with open bonds to both sites n and $n-1$. It can be shown that:

$$Q_{i+1} = p(2-p)Q_i + p(1-p)\sum_j Q_{ij} + p^2 \sum_{j+j'=1} Q_{jj'} \tag{4.30}$$

The probability Q_{i+1} that the site $n+1$ belongs to the same local cluster of size $i+1$ with the site n equals the sum of probabilities:

- The sites $n, n-1$ belong to the same local cluster of size i with probability Q_i and site $n+1$ is connected to either sites n or $n-1$ (or both) with probability $p(1-p)+p^2+p(1-p) = p(2-p)$.
- The sites $n, n-1$ belong to separate local clusters of sizes i, j respectively with probability $\sum_j Q_{ij}$ and site $n+1$ is connected to site n with an open bond with probability p and to site $n-1$ with a closed bond with probability $(1-p)$.
- The sites $n, n-1$ belong to separate local clusters of sizes j and j' where $j + j' = i$ with probability $\sum_{j+j'=i} Q_{jj'}$ and site $n+1$ is connected to both sites $n, n-1$ with an open bond with probability p^2.

We can also define $Q_{i+1,j}$ as:

$$Q_{i+1,j} = \begin{cases} (1-p)^2[Q_j + \sum_k Q_{jk}] & \text{for } i = 0 \\ p(1-p)Q_{ji} & \text{for } i \geq 1 \end{cases} \tag{4.31}$$

The probability $Q_{i+1,j}$ that sites $n+1, n$ are part of two separate local clusters of sizes $i+1$ and j equals the probabilities:

- For $i = 0$, The probability Q_{1j} that sites $n+1, n$ belong to separate local clusters of sizes $1, j$ respectively equals the sum of probabilities:
  - $(1-p)^2 Q_j$: Sites $n, n-1$ belong to the same local cluster of size j and site $n+1$ is connected to both of these sites with an open bond with probability $(1-p)^2$.
  - $(1-p)^2 \sum_k Q_{jk}$: Sites $n, n-1$ belong to separate local clusters of sizes j, k respectively and site $n+1$ is connected to both of these sites with probability $(1-p)^2$.
- For $i \geq 1$, we have the probability $p(1-p)Q_{ji}$: Sites $n, n-1$ belong to separate local clusters of sizes j, i respectively with probability Q_{ji} and site $n+1$ is connected to site $n-1$ with an open bond with probability p to form a local cluster of size $i+1$ and with a closed bond to site n with probability $(1-p)$.

To find a closed form expression for probabilities Q_i and Q_{ij} is a very difficult task. Instead we will define the following generating functions for the aforementioned probabilities:

$$H(z) = \sum_i Q_i z^i$$

$$H(z, w) = \sum_{i,j} Q_{ij} z^i w^j$$

Using these generating functions, Equations 4.30 and 4.31 give:

$$H(z, w) = z(1-p)^2[H(w) + H(w, 1)] + zp(1-p)H(w, z) \tag{4.32}$$

$$H(z) = zp(2-p)H(z) + zp(1-p)H(z, 1) + zp^2 H(z, z) \tag{4.33}$$

Each site always belongs to a cluster possibly of size 1, thus the probabilities Q_i and Q_{ij} must sum to unity $\sum_i Q_i + \sum_{ij} Q_{ij} = 1$ or equivalently:

$$H(1) + H(1, 1) = 1. \tag{4.34}$$

Finally the number of clusters of size i per lattice site N_i (density) equals the probability that a randomly chosen site is the rightmost site of such a cluster, in which case the two bonds to its right are closed with probability $(1-p)^2$. Thus the generating function for local clusters $H_0(z) = \sum_i P_0(i) z^i$ must satisfy:

$$H_0(z) = (1-p)^2[H(z) + H(z, 1)] \tag{4.35}$$

Solving Equations 4.32, 4.33 and 4.34, Equation 4.35 gives:

$$H_0(z) = \frac{z(1-p)^4(1 - 2pz + p^3(1-z)z + p^2 z^2)}{1 - 4pz + p^5(2-3z)z^2 - p^6(1-z)z^2 + p^4 z^2(1+3z) + p^2 z(4+3z) - p^3 z(1+5z+z^2)}$$

As was the case with $k = 1$ the generating function for connected clusters satisfies:

$$H(z) = H_0(z e^{2k\phi p(H(z)-1)}) \tag{4.36}$$

And the mean outbreak size is now:

$$H'(1) = \frac{H_0'(1)}{1 - 2k\phi p H_0'(1)} \tag{4.37}$$

Percolation occurs at the zero of the denominator, i.e. when $2k\phi p H_0'(1) = 1$. Thus:

$$\phi = \frac{(1-p_c)^3(1-p_c+p_c^2)}{4p_c(1+3p_c^2-3p_c^3-2p_c^4+5p_c^5-2p_c^6)} \tag{4.38}$$

where p_c is the desired percolation threshold.

The above method can be used for $k > 1$ as long as we find a way to calculate the average number of local clusters of size i. However the results will be far more complicated.

5
Simulation results

5.1 r-Harmonic Small World Model

We will study the r-Harmonic small-world model. We start with a ring lattice with n vertices and $k = 1$. Then given two vertices with labels i and j the probability for i to have j as long range contact in (R_n, ϕ_r) is given by $\phi_r(i,j) = \frac{((j-i) \bmod n)^{-r}}{H_n^r}$. Here we present a randomized algorithm for generating a network with these properties. First we create the initial structure i.e. a directed ring of n nodes. Then we choose randomly or give as input a pair of source s and target t nodes. We start at s and we randomly choose a node on the ring according to the Harmonic distribution (we give the Harmonic exponent as input). Finally we iterate until we are at a distance $O(\log n)$ where no shortcut is needed. This process is used for greedy routing on these graphs. Greedy routing is the distributed routing protocol where a node u chooses a long range contact that is closer to the target than another neighbor in order to reach the target in the minimum number of steps. It's been proven [5] that for $r = 1$ there is a tight $\Theta(\log^2 n)$ bound for the expected number of steps required for routing in the r-Harmonic ring.

r-Harmonic Distribution	Lower Bound	Upper Bound
$0 \leq r < 1$	$\Omega(n^{\frac{1-r}{2-r}})$	$O(n^{1-r})$
$r = 1$	$\Omega(\log^2 n)$	$O(\log^2 n)$
$1 < r < 2$	$\Omega(n^{\frac{r-1}{r}})$	$O(n^{r-1})$
$r = 2$	$\Omega(\sqrt{(}n))$	$O(\frac{n \log\log n}{\log n})$
$2 < r$	$\Omega(n^{\frac{r-1}{r}})$	$O(n)$

Table 5.1: Expected number of steps for greedy routing in the r-Harmonic ring

5.1.1 Computing number of hops between a source and target node

```
function ihops=hring(n, r, s, t)

%Harmonic ring:     Generate a network using a directed ring augmented with long
%                   range contacts using the r-Harmonic distribution.
%
%INPUT:             n: Number of nodes of the directed ring lattice
%                   r: Exponent of the Harmonic Distribution. For r=0 we
%                   have the uniform distribution and for r=1-log.80/log.20
%                   we obtain the Zipf distribution.
%                   s, t: Source and target nodes. If they are not given,
%                   s and t are chosen uniformly at random from the set of
%                   nodes.
%
%OUTPUT:            ihops: Number of long range contacts that are added in
%                   order to reach the target node starting from the source node.



% Initial Structure

I1=zeros(n,1);
J1=zeros(n,1);
S1=ones(n,1);

for i=1:n
    I1(i)=i;
    J1(i)=mod(i,n)+1;
end

A=sparse([I1],[J1],[S1], n, n); A=sign(A);


% Source and target nodes (s and t respectively)

if nargin <=2
   s=ceil(rand*n);
   t=ceil(rand*n);
   if (s==t)
        disp('Source node matches target node');
        return
   end
```

```
   if (mod((t-s), n)<=log(n))
        disp('Node accecible without shortcuts');
        return
   end
end

% Adding Long Range Contacts

for i=1:n
    pdist(i) =1./(i*harmonic(n,r));
end

source=s;
target=t;
dist=mod((t-s), n);
hops=0;

while dist>log(n)
   a=rand;
   B=(a>pdist);
   u=ceil(rand*length(find(B)));
   if (u~=s)
        pos=u;
        A(s,pos)=1;
        dist=mod((t-pos), n);
        s=u
        hops=hops+1;
   end

end

ihops=hops;

adj=full(A);
plotadj(adj);
```

Using MATLAB with Graphviz, we can visualize the graphs generated with this method for different values of n and r. We run the program for $n = 30$, $r = 1$ and $s = 5, t = 20$. There are needed 4 hops to reach the target node.

```
>> hring(30,1,5,20)
```

Source Node= 5
Target Node= 20
Intermediate steps=[28,29,10,19]
Number of hops = 4

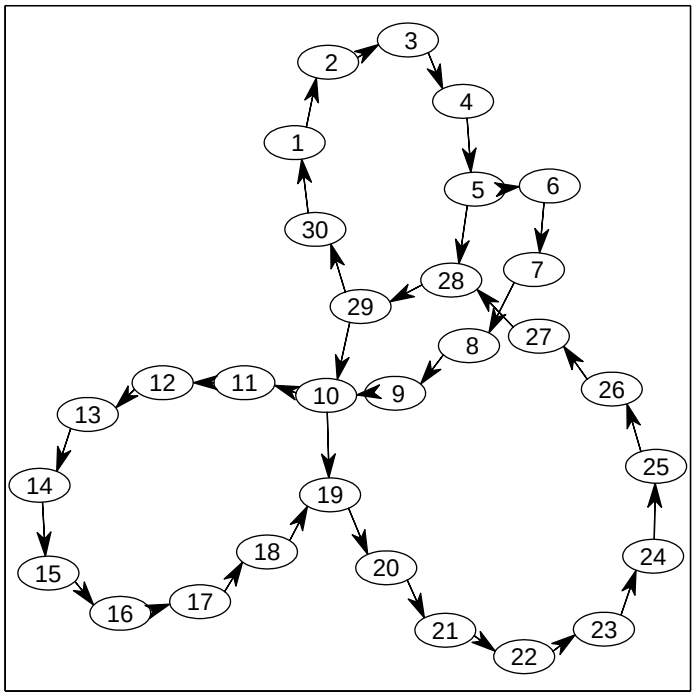


Figure 5.1: A network with $n = 20$ nodes, $r = 1$, source node=5, target node 20. Number of hops=4

5.1.2 Computing the average number of hops for a chosen source-target pair

For a given source-target pair it is interesting to see the average number of hops. We can compute this number using the function $avghops(x, n, r, s, t)$ where x is the number of repetitions.

```
function avg_hops=avghops(x,n,r,s,t)

%avg_hops:      Computes the average number of hops needed to reach a
%               target node starting from a source node.
%INPUT:         x: number of repetitions
%               n: number of nodes
%               r: Harmonic distribution exponent
%               s,t: source and target nodes
%
%OUTPUT:        avg_hops: average number of intermediate steps between
%               source and target node


if nargin <=3
   s=ceil(rand*n);
   t=ceil(rand*n);
   if (s==t)
        disp('Source node matches target node');
        return
   end

   if (mod((t-s), n)<=log(n))
        disp('Node accecible without shortcuts');
        return
   end
end


for i=1:x
    hops(i)=hring(n,r,s,t);
end

avg_hops=ceil(mean(hops));
```

Suppose we want to find the average number of hops for a network with $n = 100$ nodes, Harmonic exponent $r = 1$ and source and target nodes $s = 1$ and $t = 50$ respectively. We compute the average of 1000 different realizations of the network as:

```
>> avghops(1000,100,1,1,50)

ans = 20
```

Next we will compute the average number of hops for a network of 100 nodes, harmonic exponent $r = 1$ and source and target nodes $s = 1$ and $t = 50$ respectively:

Repetitions	No. Of Nodes	r	Source s	Target t	Average No. Of Hops
100	100	1	1	50	19
1000	100	1	1	50	20
10000	100	1	1	50	20

Based on the results, it takes on average 20 steps from source 1 to target 50. Next we will compute the average number of hops for a network of 1000 nodes, harmonic exponent $r = 1$ and source and target nodes $s = 1$ and $t = 500$ respectively:

Repetitions	No. Of Nodes	r	Source s	Target t	Average No. Of Hops
100	1000	1	1	500	146
1000	1000	1	1	500	142
10000	1000	1	1	500	145

Based on the results, it takes on average 145 steps to get to target node 500 from source node 1. Next we will change the Harmonic exponent to 1000 (approaching the geometric distribution) and try the same experiments:

Repetitions	No. Of Nodes	r	Source s	Target t	Average No. Of Hops
100	100	1000	1	50	21
1000	100	1000	1	50	20
10000	100	1000	1	50	20

Repetitions	No. Of Nodes	r	Source s	Target t	Average No. Of Hops
100	1000	1000	1	500	152
1000	1000	1000	1	500	155
10000	1000	1000	1	500	153

As it was expected it takes mores steps to route when the harmonic exponent goes to infinity. Next we will try the same experiment with harmonic exponent 1/1000 (approaching the uniform distribution):

Repetitions	No. Of Nodes	r	Source s	Target t	Average No. Of Hops
100	100	1/1000	1	50	20
1000	100	1/1000	1	50	20
10000	100	1/1000	1	50	20

Repetitions	No. Of Nodes	r	Source s	Target t	Average No. Of Hops
100	1000	1/1000	1	500	152
1000	1000	1/1000	1	500	150
10000	1000	1/1000	1	500	153

Once more as it was expected it takes more steps to route.

5.1.3 Computing number of hops between every pair of source and target nodes

Finally, we can can compute the average number of steps between every pair of source and target nodes.

```
function S=stpairs(x,n,r,s,t)

%stpairs:       Computes the average number of hops needed to reach a
%               target node starting from a source node for every source
%               and every target pair of nodes.
%INPUT:         x: number of repetitions
%               n: number of nodes
%               r: Harmonic distribution exponent
%               s,t: maximum source and target nodes
%
%OUTPUT:        S(i,j): A matrix with the average number of intermediate
%               steps between source i and target j node

for z=1:x
  for i=1:s
      for j=1:t
         S(i,j)=avghops(x,n,r,i,j);
    end
  end
end
```

Since it's not easy to present results for many values of source and target nodes, we will present a toy example of a network with 20 nodes and source and target nodes from 1 to 10:

```
>> stpairs(100, 20, 1, 10,10)

ans =

     0     0     0     7     6     7     7     8     7     6
     7     0     0     0     7     6     7     6     6     7
     7     6     0     0     0     7     6     7     7     6
     7     6     6     0     0     0     6     5     7     7
     7     6     7     7     0     0     0     7     7     8
     7     7     6     6     7     0     0     0     7     6
     7     5     6     6     7     7     0     0     0     7
     7     7     8     8     6     6     7     0     0     0
     7     7     7     6     7     8     6     7     0     0
     8     7     6     7     6     7     8     7     6     0
```

Bibliography

[1] Réka Albert, Hawoong Jeong, and Albert-László Barabási. The diameter of the world wide web. *Nature*, 401:130–131, 1999.

[2] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley, New York, 1992.

[3] Albert-Laszlo Barabási and Reka Albert. Emergence of scaling in random networks, October 21 1999. Comment: 11 pages, 2 figures.

[4] A. Barrat and M. Weigt. On the properties of small-world network models, August 25 1999. Comment: 19 pages including 15 figures, version accepted for publication in EPJ B.

[5] Lali Barrière, Pierre Fraigniaud, Evangelos Kranakis, and Danny Krizanc. Efficient routing in networks with long range contacts. In Jennifer L. Welch, editor, *DISC*, volume 2180 of *Lecture Notes in Computer Science*, pages 270–284. Springer, 2001.

[6] Marc Barthelemy and Luis A. N. Amaral. Small-world networks: Evidence for a crossover picture, March 05 1999. Comment: 5 pages, 5 postscript figures (1 in color), Latex/Revtex/multicols/epsf. Accepted for publication in Physical Review Letters.

[7] B. Bollobás. *Random Graphs*. Academic Press, 1985.

[8] Reinhard Diestel. *Graph Theory*. Springer-Verlag, New York, 2 edition, 2000.

[9] P. Erdos and A. Rényi. On random graphs. *Publ. Math. Debrecen*, 6:290–291, 1959.

[10] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5:17–61, 1960. A seminal paper on random graphs. Reprinted in *Paul Erdős: The Art of Counting. Selected Writings*, J.H. Spencer, Ed., Vol. 5 of the series *Mathematicians of Our Time*, MIT Press, 1973, pp. 574–617.

[11] Jerrold W. Grossman. Paul Erdös: The master of collaboration. In Ronald L. Graham and Jaroslav Neštřil, editors, *The Mathematics of Paul Erdös II*, pages 467–475. Springer-Verlag, Berlin, 1997.

[12] Rajesh Kasturirangan. Multiple scales in small-world networks. Technical Report AIM-1663, MIT Artificial Intelligence Laboratory, August 11 1999.

[13] J. Kleinberg. The small-world phenomenon: an algorithmic perspective. In *Proceedings of the 32nd ACM Symposium on the Theory of Computing*. 2000.

[14] C. Korte and S. Milgram. Acquaintance networks between racial groups: Application of the small world method. *J. Personality and Social Psych.*, 15:101, 1978.

[15] Albert laszlo Barabási, Reka Albert, and Hawoong Jeong. Scale-free characteristics of random networks: The topology of the world-wide web, September 25 2000.

[16] S. Milgram. The small world problem. *Psychology Today*, 1(1):60–67, 1967.

[17] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. 2002. Technical Report.

[18] Cristopher Moore and M. E. J. Newman. Epidemics and percolation in small-world networks, January 06 1999. Comment: 6 pages, including 3 postscript figures.

[19] Cristopher Moore and M. E. J. Newman. Exact solution of site and bond percolation on small-world networks, January 26 2000. Comment: 13 pages, 3 figures.

[20] Rajeev Motwani and Prabhakar Raghavan. Randomized algorithms. *ACM Computing Surveys*, 28(1):33–37, March 1996.

[21] M. E. J. Newman. Models of the small world: A review, May 09 2000. Comment: 9 pages including 3 postscript figures, bibliography updated and minor corrections to text in this version.

[22] M. E. J. Newman. The spread of epidemic disease on networks. *Physical Review E*, 66:016128, 2002.

[23] M. E. J. Newman. Random graphs as models of networks. In *Handbook of Graphs and Networks: From the Genome to the Internet*, pages 35–68, Weinheim, 2003. Wiley-VCH Verlag.

[24] M. E. J. Newman. The structure and function of complex networks. *SIREV: SIAM Review*, 45, 2003.

[25] M. E. J. Newman, C. Moore, and D. J. Watts. Mean-field solution of the small-world network model, September 21 1999. Comment: 14 pages, 2 postscript figures.

[26] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64:026118, 2001.

[27] M. E. J. Newman and D. J. Watts. Scaling and percolation in the small-world network model, May 06 1999. Comment: 12 pages including 9 postscript figures, minor corrections and additions made in this version.

[28] Joel Spencer. *Ten Lectures on the Probabilistic Method.* Regional Conference Series on Applied Mathematics (No. 52). SIAM, 1987.

[29] Dietrich Stauffer and Ammon Aharony. *Introduction To Percolation Theory.* CRC, July 1994.

[30] S. Wasserman and K. Faust. *Social Network Analysis.* Cambridge University Press, Cambridge, 1994.

[31] D. J. Watts. *Small Worlds.* Princeton University Press, Princeton, 1999.

[32] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):397–498, June 1998.

[33] Herbert S. Wilf. *Generatingfunctionology*. A. K. Peters, Ltd., Natick, MA, USA, 2006.

Index