

Advanced Modeling in Biology Assignment

Small-World Networks: Characteristics and Consequences

Introduction & Background

Networks form a large part of our life, ranging from the transportation infrastructure to the internet, both exemplifying the necessity to study and understand network theory. Furthermore, the study of dynamical networks such as coupled oscillators and metabolic networks can help us to understand and predict the behavior of such networks with implications in fields such as laser development and pharmaceutical design. For now, we remain on the topic of simple static networks as dynamic networks are inherently difficult to understand. Complications such as structural complexity, network evolution, connection type diversity, dynamical complexity, node diversity, and various other intricacies can influence each other and make the study of such dynamical systems difficult. (Strogatz, 2001)

Watts and Strogatz in their 1998 letter to *Nature* organized networks into three different categories based upon two quantifiable measures: the characteristic path length (L) and the clustering coefficient (C). Regular networks, where vertices of the graph are connected in a specifically defined way, show long path lengths and high clustering coefficients. On the opposite side of the spectrum, random graphs, where vertices are randomly connected, show both short path lengths and low clustering coefficients. In the middle, there is a unique region where high clustering coefficients and low path lengths dominate. These networks were coined 'small-world' networks after the six-degrees of freedom phenomenon.

Watts and Strogatz also formulated a systematic way to produce these 'small-world' networks by first taking a regular graph, selecting an edge at random, and with some probability p , rewiring it to attach to a different vertex (see figure 1). The authors then went on to explore the implications of such graphs, especially the small-world graphs compared to both regular and random graphs.

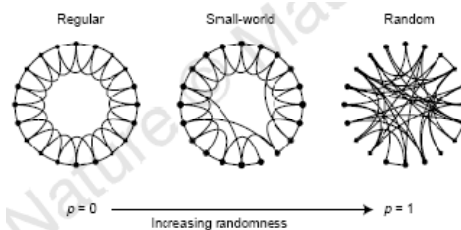


Figure 1. Random rewiring procedure produces a random graph as p approaches 1 starting with a regular graph. Reproduced from Watts & Strogatz, 1998.

Main Results of Paper

The main results of Watts and Strogatz's 1998 letter to *Nature* are two measures by which we can characterize unweighted networks into the three different types of graphs. They began with a simple ring lattice with n vertices and k edges per vertex. To quantify their results, they defined the characteristic path length L as the average distance from one vertex to another vertex, normalized to the characteristic path length for the completely regular graph. Mathematically, this can be represented in equation 1 below where d_{ij} is the matrix with the shortest path lengths between two vertices i and j . From this connection matrix, either Dijkstra's algorithm or the Floyd-Warshall algorithm can be used to calculate the minimum distance between two points to form our distance matrix d_{ij} . Both algorithms for calculating the minimum distances are non-trivial and are of polynomial order (Latora & Marchiori, 2001). Hence, for large graphs, these algorithms become computationally intensive.

$$L = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij} \quad (1)$$

Newman, Moore, and Watts further developed analytical solutions to the characteristic path length in their 2000 paper. Furthermore, their solution takes into account different numbers of vertices and shortcuts, one thing Watts & Strogatz's paper does not address.

Watts and Strogatz also defined the clustering coefficient (C) being the sum of the number of edges existing in a subgraph of neighbors (C_i) also normalized to the maximum possible number of neighbors (equation 2) (Latora & Marchiori, 2001). In terms of social networks, the clustering coefficient is easy to understand. If two people are friends, both of them are likely to have another friend in common. From a biological perspective (e.g. protein-protein interactions), proteins which interact tend to also share interactions with other proteins.

$$C = \frac{1}{n} \sum_i C_i \quad (2)$$

Watts and Strogatz's results from their theoretical rewiring experiment are shown in figure 2. Small-world networks are defined arbitrarily to be those networks found between $0.001 < p < 0.1$, exhibiting high clustering coefficient but relatively low path lengths.

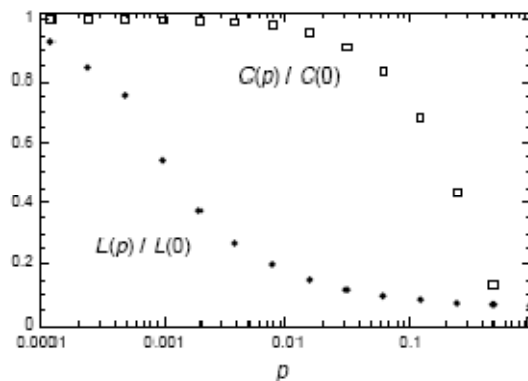


Figure 2. Characteristic path length and clustering coefficient normalized with respect to regular graphs for various rewiring probabilities averaged over 20 repeats. For these graphs, $n = 1000$ vertices with average degree of $k = 10$ edges per vertex. Reproduced from Watts & Strogatz, 1998.

With these results in mind, Watts & Strogatz investigated three different networks and concluded that all three are indeed small-world networks by comparing both the characteristic path lengths and clustering coefficients to randomly generated graphs with the same number of vertices and edges. They also used the spread of an infectious disease to investigate the functional significance a small-world network. They saw that very few short cuts are required to cause a significant increase in the spread rate of a disease.

Applications of Small-World Networks

The most obvious application of small-world networks is in immunizations and controlling epidemics as explored briefly in Watts & Strogatz's paper and in more depth in Pandit & Amritkar's 1999 paper on the characterization and control of small-world networks. Through various theoretical experiments on immunizing different nodes and investigating the spread an epidemic, they concluded that by selective immunization of nodes along far edges (short cuts), they were able to most effectively control and stop the epidemic.

A more recent application has been in systems biology researching into protein-protein interactions and other metabolic networks (Zhu et al., 2007). Biology has inherently been complex but certain characteristic pathways and interactions are highly conserved through all living organisms. Analyzing these networks can help us to understand some of the basic mechanisms controlling cellular processes and the spread of disease within an organism itself. Furthermore, as the structure of a network usually determines its function, there has been a recent drive to find network motifs in interaction networks in the quest to uncover the functions of several uncharacterized proteins.

These are only a couple of examples of small-world network applications. Further applications have been found in food webs, power grid networks, and even theoretical physics (Strogatz, 2001). In fact, most networks found have been shown to be small-world networks.

Discussion & Conclusion

Although the paper presented the information well, there is unfortunately no indication as to the method they used to calculate the shortest path between two points. We are left to study the methods of another paper (Latora & Marchiori, 2001) to decipher their methods. As mentioned before, this can be computationally difficult, especially for a large number of points. In the code used to recreate the results, the Floyd-Warshall

method of finding the shortest path length between two points was used. Since this algorithm goes through every possible scenario, the minimum is guaranteed to be found. However, for networks that are very large such as the internet, using such an algorithm would be relatively inefficient without significant computational resources.

To obtain a better understanding of their methodology, the rewiring procedure was attempted on a small scale network with $n = 100$ and $k = 8$. The results shown in figure 3 are somewhat similar to those obtained in Watts & Strogatz's paper, but show that the transition between regular to small-world to random occurs at higher probabilities of rewiring (leftward shift of the curve).

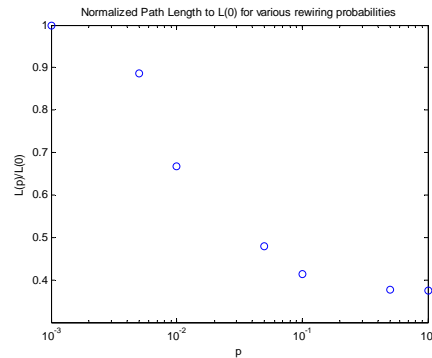


Figure 3. Matlab simulated characteristic path length normalized with respect to regular graphs for various rewiring probabilities. For these graphs, $n = 100$ vertices with average degree of $k = 8$ edges per vertex. Slow run time for the application was the major factor preventing a larger scale simulation with more vertices. These results are in line with Newman, Moore and Watts' mean field solution with the number of vertices normalized with the number of shortcuts (Newman, et al., 1999).

Furthermore, as seen in our recreation of results, the probability bounds in which the small-world phenomenon occurs also depends upon the number of vertices and edges the graph contains. The analytical solution developed by Newman, et al., also takes into account variations in the number of vertices and shortcuts. However, the overly simple definition of a small-world network that Watts & Strogatz developed is too broad and tends to encompass almost any network which is neither fully regular nor totally random. Further extensions to this paper might be to investigate the differences between small world networks with a penchant towards the random side and towards the regular side.

Recent studies by Latora and Marchiori (2001) have introduced another network measure than the characteristics defined by Watts & Strogatz. Path length and clustering coefficient might not be the best characteristics to look for in a network, but instead, one should consider the efficiency of a network, defined as being the inversely proportional to the shortest distance between two vertices. Small-world networks are defined to be both globally and locally efficient, setting them apart from both regular (only locally efficient) and random (only globally efficient) graphs. Although this reduces the computational task of determining which type of network one is looking at, this again does not differentiate networks within the small-network category.

One other aspect of networks that Watts & Strogatz did not go into detail is the importance of either far away connections (short cuts) compared to the importance of hubs. Pandit and Amritkar have argued that it is the existence of far away connections that contribute to the efficiency of small-world networks. Indeed, following their theoretical experiment on immunizations and controlling epidemics, it was shown that having a few short cuts drastically increases the spread of an epidemic, while immunizing large hubs did little to combat this efficient spread of disease. In contrast, Zhu et al. have argued that hubs are more important in biological networks, especially when considering protein-protein interactions since they are more likely to be essential proteins and evolutionarily conserved relative to non-hub proteins. It is clear that an epidemic must be approached differently than the proteomic network and one needs to be wary of placing importance on certain characteristics before understanding the consequences.

Overall, Watts & Strogatz's 1998 paper paved the way to investigate networks and coming up with basic quantitative measures to identify types of networks. Further research since 1998 has indicated increasing realization of the importance of networks, most recently within the field of systems biology. With an increasing desire to be able to control genetic and metabolic expression, it is important to understand the transfer of information through networks to successfully intervene and alter such pathways to our desire.

References

Floyd–Warshall algorithm. 2008. In *Wikipedia, The Free Encyclopedia*. Retrieved 15:30, April 17, 2008, from http://en.wikipedia.org/w/index.php?title=Floyd%E2%80%93Warshall_algorithm&oldid=205000013

Latora, M. and Marchiori, M. 2001. Efficient Behavior of Small-World Networks. *Physical Review Letters* **87**(19):198701.

Newman, M., Moore, C., and Watts., D. 2000. Mean-Field Solution of the Small-World Network Model. *Physical Review Letters* **84**(14): 3201-3204.

Pandit, S. and Amritkar, R. 1999. Characterization and control of small-world networks. *Physical Review E* **60**(2): 1119-1122.

Strogatz, S. 2001. Exploring Complex Networks. *Nature* **410**: 268-276.

Watts, D. and Strogatz, S. 1998. Collective dynamics of ‘small-world’ networks. *Nature* **393**: 440-442.

Zhu, X., Gerstein, M., and Snyder, M. 2007. Getting Connected: analysis and principles of biological networks. *Genes & Dev* **21**: 1010-1024.

Appendix A: Matlab Code

Below is the Matlab code for creating a regular network and running the rewiring procedure. The program also calculates the normalized characteristic path length of the network.

```
%Creating and exploring networks

clear all; clc; close all;

n = 100; %Total number of vertices
k = 8; %Number of connections from each vertex (must be even)

%Formation of a 'regular' graph
distmat = zeros(n);
for i = 1:n-(k/2)
    for j = 1:k/2
        distmat(i,i+j) = 1;
        distmat(i+j,i) = 1;
        if i < k
            distmat(i,n+1-j) = 1;
            distmat(n+1-j,i) = 1;
        end
    end
end

%Floyd-Marshall Algorithm for finding shortest distance between two points
%Initialize matrix for path lengths
for i = 1:n
    for j = 1:n
        if distmat(i,j) == 0
            newdist(i,j) = 1000; %Simulate infinity
        else
            newdist(i,j) = distmat(i,j);
        end
    end
end

path = newdist;

for k = 1:n
    for i = 1:n
        for j = 1:n
            path(i,j) = min([path(i,j) path(i,k)+path(k,j)]);
            if i == j
                path(i,j) = 0;
            end
        end
    end
end

%Finding the L value (characteristic path length)
%Path length given by formula 1/N(N-1) * sum(d(ij))

L0 = (1/(n*(n-1)))*sum(sum(path));

rewireprobs = [0.001 0.005 0.01 0.05 0.1 0.5 1];
repeat = 7;

storevals = zeros(length(rewireprobs),repeat);

%Probability of rewiring
for c = 1:length(rewireprobs)
    for f = 1:repeat
        rewireprob = rewireprobs(c);
        rewiretimes = 2000;
        counter = 0;
        %Random rewiring
        for times = 1:rewiretimes
            randEdgeToRewireFrom = [ceil(rand*n) ceil(rand*n)];
            randVertexToRewireTo = ceil(rand*n);
            %Make sure that there is an edge to break
            test2 = distmat(randEdgeToRewireFrom(1),randEdgeToRewireFrom(2));
```

```

%Make sure not making another edge already there
test3 = distmat(randEdgeToRewireFrom(1),randVertexToRewireTo);
if (rand < rewireprob) && (test2 == 1) && (test3 == 0)
    distmat(randEdgeToRewireFrom(1),randEdgeToRewireFrom(2)) = 0;
    distmat(randEdgeToRewireFrom(2),randEdgeToRewireFrom(1)) = 0;
    distmat(randEdgeToRewireFrom(1),randVertexToRewireTo) = 1;
    distmat(randVertexToRewireTo,randEdgeToRewireFrom(1)) = 1;
    counter = counter + 1;
end
end

%Floyd Marshall again
%Initialize matrix for path lengths
for i = 1:n
    for j = 1:n
        if distmat(i,j) == 0
            newdist(i,j) = 1000; %Simulate infinity
        else newdist(i,j) = distmat(i,j);
        end
    end
end

alterpath = newdist;

for k = 1:n
    for i = 1:n
        for j = 1:n
            alterpath(i,j) = min([alterpath(i,j) alterpath(i,k)+alterpath(k,j)]);
            if i == j
                alterpath(i,j) = 0;
            end
        end
    end
end

%Find the new length and ratio

Lnew = (1/(n*(n-1)))*sum(sum(alterpath));
Lratio = Lnew/L0;
storevals(c,f) = Lratio;
end

for m = 1:length(rewireprobvals)
    avgvals(m) = mean(storevals(m,:));
end
figure;
semilogx(rewireprobvals,avgvals, 'o')
title('Normalized Path Length to L(0) for various rewiring probabilities')
xlabel('p')
ylabel('L(p)/L(0)')

```