


Chapter 3

Small World Networks

3.1 Background

Networks are often characterized by commonly studied properties of their nodes and edges. Two of these properties are called the clustering coefficient and characteristic path length. As discussed in Watts and Strogatz (1998), the clustering coefficient depends on triplets of nodes. Consider a vertex V with k_V neighbors on an undirected graph. Any connection between the neighbors of V generates a node triplet. Then there can exist at most $k_V(k_V - 1)/2$ edges between the k_V neighbors. Let clustering coefficient C_V denote the fraction of these triplets that actually exist for node V . Then, the average clustering coefficient C is the average of C_V calculated from each vertex. The characteristic path length, L of a graph is the shortest path length between two vertices, averaged over all unique pairs of vertices.

Historically, many real world networks have been classified as mostly regular or random. Regular networks are defined as networks where each node has the same number of connections. Random networks are defined wherein the connections between nodes are determined randomly. Small world networks were introduced by Watts and Strogatz (1998) as a sort of middle ground between the two. It is a common property of regular

networks, to be highly clustered. Random graphs, on the other hand, exhibit lower clustering but have small characteristic path lengths due to the existence of many randomly introduced connections between distant nodes (Walsh et al., 1999). 

We are interested in small world networks because they could be analogous to the movement network of invasive species. Population networks of invasive species often develop by a combination of movement between neighboring habitats and dispersal along a long distance transport vector, for example, a truck carrying a soil shipment across the United States. Indeed, there have been many research papers discussing small-world properties of real world networks, including transportation networks (Latora and Marchiori, 2001). By analyzing the impact of randomness on a small world network, we hope to learn about the effect of long-distance dispersal in biotic invasions.

3.2 Small World Methods

The original paper (Watts and Strogatz, 1998) generates a small world network by rewiring a regular ring lattice into a random graph. We begin with the regular ring lattice, which we define as a regular network where the nodes are arranged into a ring, and begin randomly choosing edges and randomly reassigning their endpoints, based on a parameter of rewiring probability p . An example of this process is shown in Figure 3.1.

The increase in rewiring probability is correlated with a decline in characteristic path length and mean clustering coefficient, seen in Figure 3.2.

3.2.1 Extension to Invasion Model

In this section we implement a portion of the procedures including generating the small world network in R using the package `igraph` (Csardi and Nepusz, 2006).

The algorithm for constructing the small world network using `igraph` begins by constructing a ring lattice with each of the n nodes connected to its nearest $2m$ neighbors.

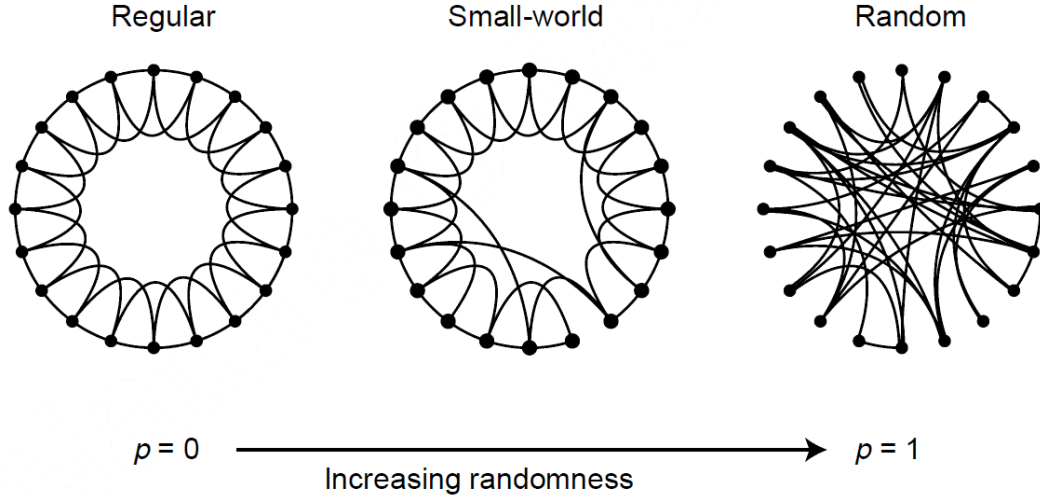


Figure 3.1: A regular ring lattice is rewired with increasing values of p , the rewiring probability.

We choose $m = 2$ and $n = 50$. Begin the rewiring process by choosing nodes x and y from a random uniform probability distribution of all possible node values. Rewiring is done by replacing each edge e_{ij} with edge e_{xy} with probability p . Rewiring events resulting in a self-loop or a duplicated path are redrawn by the igraph algorithm, by redrawing another pair of values x and y . We imposed an additional check at the end of the rewiring process so that ~~we entirely reconstructed~~ any disconnected networks, meaning networks where there are unreachable nodes. ~~By reconstructing we mean restarting the rewiring process entirely from the original ring lattice configuration.~~ This implementation slightly differs from the original Watts and Strogatz method in that both endpoints of the edge e_{ij} are randomized here as opposed to just one in the original method. However, this difference did not seem significant to the underlying small world network properties. To validate the igraph method we created a plot of characteristic path length and clustering coefficient in Figure 3.2 and **compared the results** with those presented in Watts and Strogatz (1998).

To adapt this small world network to fit our model of species invasion, we transform its

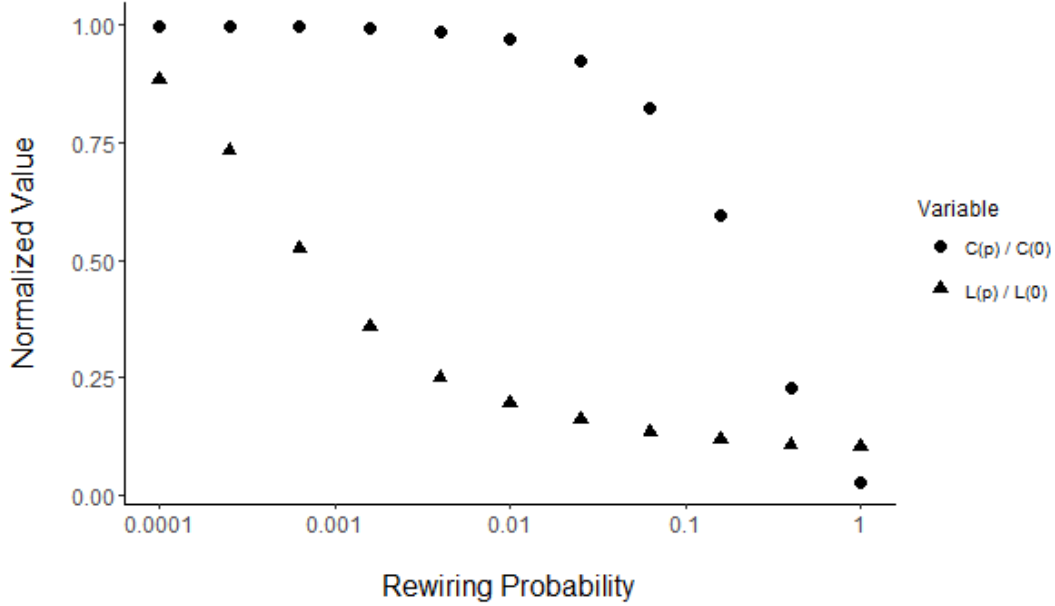


Figure 3.2: Plots showing relationship between normalized characteristic path length and clustering coefficient with rewiring probability. $C(p)$ and $L(p)$ refer to the values of each property for a network generated with given rewiring probability p . The plotted values are normalized by their values calculated on a regular graph. Parameter values are $n = 1000$, $m = 5$, the results are averaged over 20 realizations.

adjacency matrix. The adjacency matrix of a network is defined as a matrix \mathbf{A} such that every element a_{ij} has a binary value 1 or 0. A value of 1 indicates a connection between nodes i and j whereas 0 indicates no connection. Because the small world network is undirected, the adjacency matrix is symmetric, so that $a_{ij} = a_{ji}$.

To map the adjacency matrix to the transition matrix, we perform the following transformations:

1. $\mathbf{P}_{ij} = va_{ij}$ for $i \neq j$
2. $p_{jj} = 1 - \sum_{i=0}^n va_{ij}$

In this definition, \mathbf{P} refers to the transition matrix defined in Equation 2.8, and v is the

migration rate, the proportion of population sent from node p_{ij} to each other node.

Consider the network in Figure 3.3 of 50 nodes labeled from 1 to 50 counterclockwise in a ring lattice. We use the transformed adjacency network to map an invasion with yearly time steps over the course of 1000 years. We investigate the transient properties of this system by recording a metric $t(p)$, called the time to establishment. We define this metric on a network rewired with probability p , as the number of time steps it takes for the population to spread from the initial node 1 to the one diametrically opposite, node 26, and surpass a population threshold arbitrarily set to 100 individuals.

Consider one experiment to be defined as generating a small world network with rewiring probability p , transforming the adjacency matrix, and running a simulation as described above.

We performed several sets of experiments recording $t(p)$ at different values of p and averaged the results over 1000 networks for each value of p , with the average times to establishment denoted by $T(p)$. In each set of experiments, we varied values of v and r to observe the effects of migration rate and birth rate on the transient properties an invasion on our network model. The results are presented in the next section.

3.3 Results

We first present the results of normalized time to establishment, that is, $T(p)/T(0)$, versus the rewiring probability p , with varying migration rates v in Figure 3.4. We next present the results of varying the birth rate r in Figure 3.5. Then, we present results where v and r are varied respectively to each other. The migration and birth rate terms are both linear order in population N , so varying both at the same time in opposing directions attempts to show the relative effects of each parameter while keeping the overall time scale of the invasion reasonably constant. The results of the normalized time to establishment for this set of experiments is shown in Figure 3.6, and the non-normalized, absolute time ~~valued~~ results are presented in Figure 3.7.

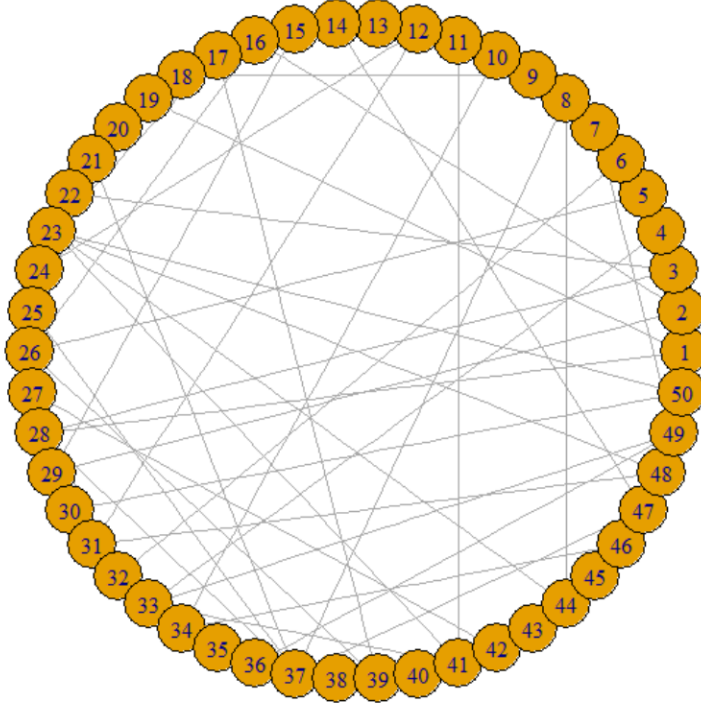


Figure 3.3: Example of the **smallworld** network generated by igraph with $n = 50$, $m = 2$, and $p = 0.25$.

Comparing the plots of normalized time to establishment to Figure 3.2, we find that the normalized time to establishment, $T(p)/T(0)$, shares the same shape as $C(p)/C(0)$. This is interesting because we expected these paths to relate to the characteristic path length since **it** is directly related to the number of long-distance connections between distant nodes on the ring. It is not immediately clear why the time to establishment should appear more analogous to the clustering coefficient than the characteristic path length, but **it** warrants further investigation.

Following the **paths** in ~~the~~ Figure 3.4, we find that starting from $p = 0$, the paths diverge and settle at $p = 1$ **in order of the magnitude of the migration rate**. The larger v

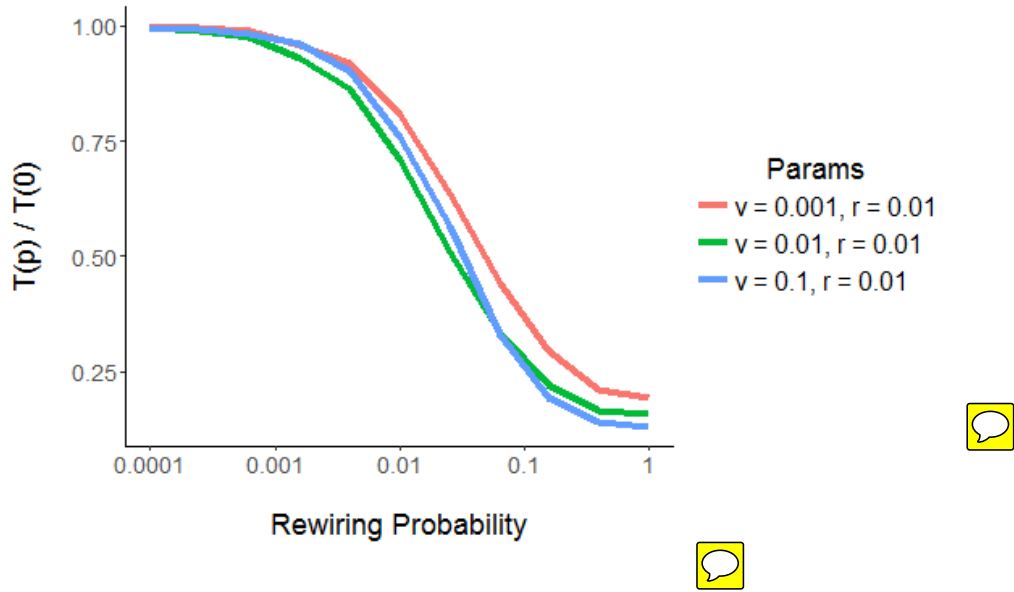


Figure 3.4: Plot of the normalized time in years to establishment of a species invading a small world network with varying values of v . Parameters $n = 50$, $m = 2$, $r = 0.01$, and establishment threshold is set at 100.

is the smaller $T(1)/T(0)$. This is as expected as we hypothesized a relationship between decreased characteristic path length and increasing speed of invasion (time to establishment). This is reasonable, as in a highly connected graph with high migration rate, the first few time steps could find the invasive species spread all around the network, whereas this would not occur as quickly with a slower migration rate.

Following the paths in the Figure 3.5, we find a surprising result. First, we notice that whereas previously observed paths were fairly close together, at very low birth rate, the time to establishment drops off at a higher value of p than at higher birth rates. Secondly, we notice that the order of convergence is inversely related to the birth rate. Note that we are plotting the normalized time to establishment. The absolute time to establishment is positively related to increasing values of both v and r . The fact that the normalized time to establishment at $p = 1$ is inversely related to birth rate, such that the path with the lowest growth rate exhibits the lowest normalized time to establishment at $p = 1$ can be

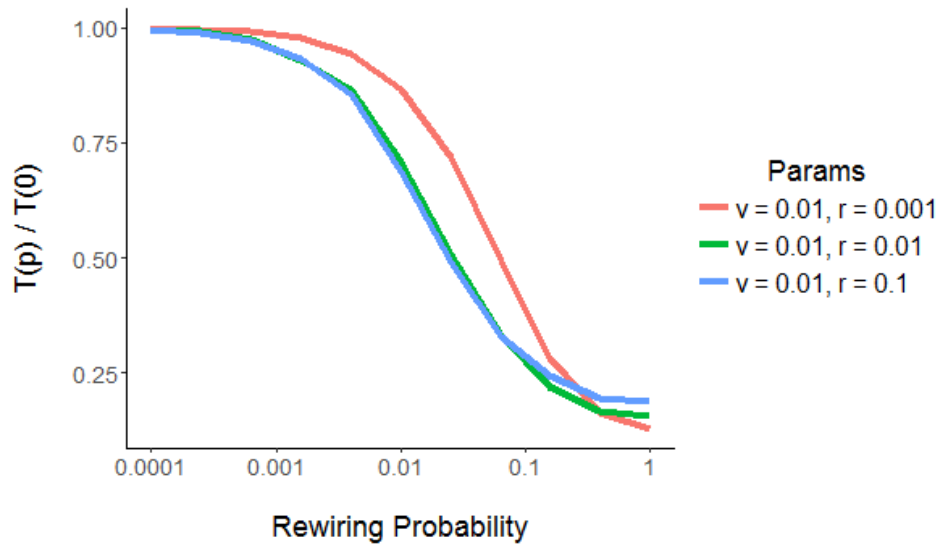


Figure 3.5: Plot of the normalized time in years to establishment of a species invading a small world network with varying values of r . Parameters $n = 50$, $m = 2$, $v = 0.01$, and establishment threshold is set at 100.

explained by considering the contribution of long distance connections in the spread of the invasion. One possible explanation is that the long distance connections contribute more in invasions with low birth rate because the time it takes for an invasion to spread via diffusion across the network would be much longer than for an invasion with a higher growth rate. Thus, at low values of p , the invasion process is extremely slow with low birth rate, but increases rapidly with an increasing number of long distance connections. This could also explain the large gap between the paths. Perhaps this phenomenon is only visible at sufficiently low birth rate.

By varying both parameters v and r concurrently, in Figure 3.6, we see that the order of the time to establishment at $p = 1$ supports the results from varying the parameters separately. However, if we compare this to the paths of charting absolute time to establishment in Figure 3.7, that is, the number of time steps, we notice two main differences. First, we notice that the values of $T(p)/T(0)$ between the paths diverge as p increases,

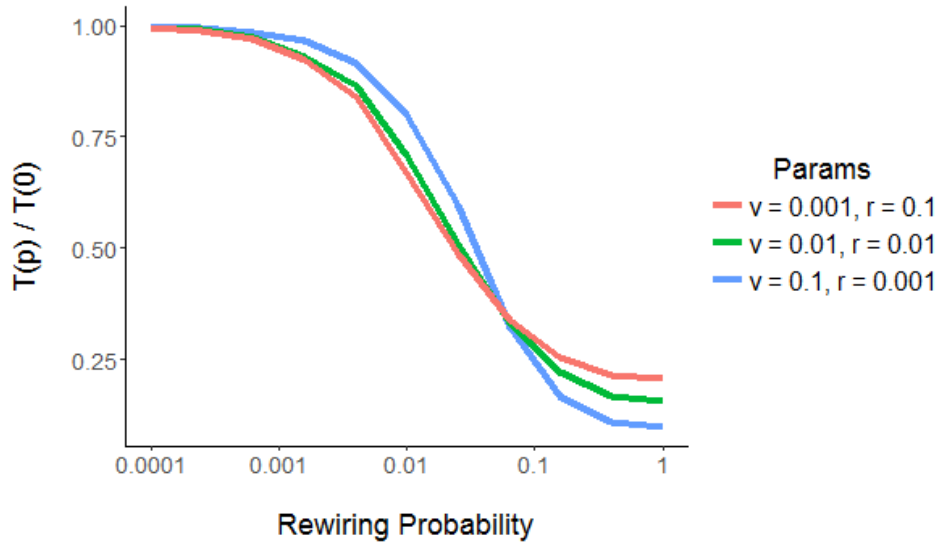


Figure 3.6: Plot of the normalized time in years to establishment of a species invading a small world network with varying sets of migration rates v and birth rate r . Parameters $n = 50$, $m = 2$, $r = 0.01$, and establishment threshold is set at 100.

the absolute time values of $T(p)$ converge as p increases. One possible reason for this is that in the initial regular graph, the shortest path **to between** opposite nodes is longer than in the random graph. Different combinations of parameters v and r could produce slower or faster spreading invasions. As that path length decreases as p increases, any difference in the spread rate between parameter combinations is diminished since there is less distance to cover.

Additionally, we find that the ~~curves tracing~~ absolute time to establishment vs. p **converge** in a different order at $p = 1$ than those of normalized time to establishment. Here, the curve with the intermediate value of v and the lowest value have switched places. It seems that in absolute terms, the curve with $v = 0.1$ still results in the fastest establishment. There is possibly some interplay between these parameter values and invasion dynamics that we do not completely understand.

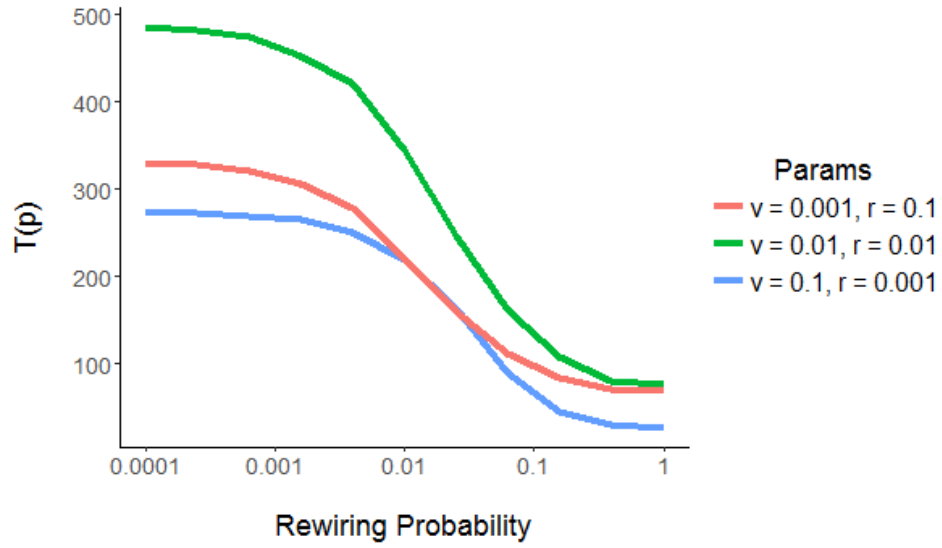


Figure 3.7: Plot of the absolute time in years to establishment of a species invading a small world network with varying sets of migration rates v and birth rate r . Parameters $n = 50$, $m = 2$, $r = 0.01$, and establishment threshold is set at 100.