Percolation on the Kleinberg Small-World Networks

Introduction

In this project, I determined the site percolation threshold of the Kleinberg small-world model (Kleinberg 2000). Site percolation on a lattice consists of labeling each site on the lattice "active" or "inactive." Percolation can be used to model the spread of epidemics and catalytic surfaces. Applying site percolation to small-world networks could be used to better understand the spread of diseases and innovation. The Kleinberg small-world network (2000) is similar to a two-dimensional version of the Watts-Strogatz small-world model. The main difference between the two models, besides the dimensionality, is that the long-range connections in the Kleinberg model are directed. The main feature that Kleinberg wanted to show is that there is an optimal value to the clustering coefficient, r, for greedy search with local information. The clustering coefficient determines the probability that a long-range bond, say from a to b, is chosen, $P(a \rightarrow b) \propto (|a_x - b_x| + |a_y - b_y|)^{-r} = d(a,b)^{-r}$, where $d(a,b)$ is the Manhattan distance between sites a and b. Note that his definition of clustering coefficient is almost completely unrelated to the typical definitions of clustering coefficient. Kleinberg stated that his model is closer to reality than other small world network models because it produces similar paths for greedy local search as the original Milgram experiment. I used his model to begin to answer questions about message passing when some of the people between the source and the target, in this fictitious version of Milgram's experiment, always throw the letter in the trash.

Motivation:

Percolation on small-world networks can answer questions about the spread of disease through networks that resemble real human contact networks. In this context the 'active' nodes are people who are susceptible to the disease; 'inactive' nodes are people who are immune to disease. By determining the effect of different long-distance contact distributions and number of long-distance contacts on the percolation threshold, we can start to answer epidemic prevention questions. One possible question is, "is it more important to reduce the total number of long-range contacts or changing the distribution of long range contacts to make the contacts more 'clustered'?"

Another motivation is my initial motivation: modeling the diffusion of information through high-loss small-world networks. One known issue with Milgram's experiment is that a number of the messages were dropped by intermediate people. In terms of percolation these people would be 'inactive' nodes. With both stochastic and deterministic message passing schemes, a message that goes through more steps is more likely to be dropped. Therefore, the average path length of a message that makes it to its destination is shorter than the average shortest path if there were no people who dropped messages. I planned to look into the probability of messages reaching a target at the percolation threshold. As it turned out, just finding the percolation threshold as a function of r and the total number of long connections (N) was difficult enough.

Related work:

The subject of percolation is very old and the percolation threshold has been determined for practically all 2D (Suding 1999) and 3D (Lorenz 2000) lattices. Lattices are said to percolate if there exists a path from one side of the lattice to the other through only active sites. This group of active sites is called a percolation cluster. Networks, on the other hand, require a different definition of percolation. Percolation on networks requires determining the size of the giant component of the network, after the sites have been labeled active or inactive. Network percolation has been performed on random graphs (Callaway 2000), 1D Watts-Strogatz small-world networks (Newman 1999), 2D Watts-Strogatz small-world networks (Newman 2002), and on Heterogeneous networks (Sander 2002). Due to the similarity between the Kleinberg and Watts-Strogatz models, one *limit* of my analysis should yield similar results to Newman 2002.

Methods:
My goal is to determine the percolation threshold as function of the two parameters of the Kleinberg model, the power-law distribution coefficient (or 'clustering coefficient') r, and the number of long range connections N. The percolation threshold is defined as the percentage of active sites required to always percolate for infinite size systems.

To find the percolation threshold I used simulations written in C++. The details of my simulations are as follows. I create a square lattice of equal height and width. Each site gets N additional directed long range links. If N is not an integer, then some sites get $\lfloor N \rfloor$ and the others get one more ($\lceil N \rceil$) such that the average number sites is made to be N. The links are picked such that the probability of a link of a given length is given by d^{-r}. After the long links are added, the lattice is percolated: p sites are made active and (1-p) sites are made inactive. The last step is to determine if there is a path from any of the active sites on the bottom (if any exist) to the active sites on the top (if any exist) by only hopping on active sites. Another way to say this is that I looked for a spanning cluster. If such a cluster exists, I said the system percolates. I averaged over a large number of different long-link configurations are percolated lattices to find the average probability of the system percolating, $\Pi(p)$.

I measured the profiles of $\Pi(p)$ for several size systems, Figure 1. The most common method of determining the percolation threshold is to measure p_c, say where $\Pi(p) = 0.5$, for different systems and use finite size scaling to extrapolate to infinite size systems. Specifically, the finite size corrections are known to have a power-law form, Figure 2.

Results:
My main result is shown in Figure 3. This figure shows how p_c varies as a function of r and N. There is a lot to digest in the figure so I will discuss some of the interesting behavior. The first limit to note is the limit of N = 0; this corresponds to a normal square lattice with no long-range connections. The percolation threshold for the square lattice is known to be 0.59274621(13). My simulation calculated p_c to be 0.5927. This limit shows, at least partial, functionality of my programs.

The r = 0 Limit

This limit is shown in Figure 4. This limit corresponds to long bonds of any length being allowed. This result can be compared with the results of Newman 2002. For both results we see the percolation threshold start at that of the square lattice, 0.5 and 0.5927 for bond and site percolation, respectively. For my results, p_c monotonically decreases until N=2, after which p_c is a constant value, 0.1897. Newman's results don't show this behavior and if I understand their paper correctly, they don't predict any point where p_c should become constant. This discrepancy is because, in that paper, he did not measure p_c for N>1 and he used the giant component to determine percolation, whereas I used crossing probability.

The r = ∞ Limit

My approximation of this limit is shown in Figure 5. The slice of the contour plot is at r=15, a good approximation of the infinite behavior. From the plot we can see that there is a linear decrease in p_c until N=2 where p_c becomes constant. One interesting thing about this limit is that it can be compared against other lattices because it corresponds to adding some connections to next-nearest neighbors. In terms of lattices, a square lattice has a coordination number (z) of 4, i.e. four neighbors. N=1 and N=2 correspond to z=5 and z=6, respectively. I found that p_c for z=5 and z=6 are 0.528 and 0.470. These results can be compared against the percolation threshold of normal lattices. My results for z=5 are smaller than known lattices (ranging from 0.550 to 0.579). The common z=6 lattice is the triangular lattice which has a site percolation threshold of 0.5. In all of these cases, the r=∞ limit have a lower percolation threshold than normal lattices. This is an interesting result, without a clear explanation.

The N=1 and N=2 Limits

A plot of p_c versus r for N=1 and N=2 is shown in Figure 6. For large r, p_c approaches some asymptotic value as the 'long' links become next-nearest neighbor links, as seen in Figure 5. When r decreases, p_c also decreases until it reaches a constant value at r ≈1.5. Why should p_c stop decreasing a function of r? For very small r there is a non-zero chance of making a connection directly across the lattice from start to finish. The source of this behavior is my definition of finite p_c. Figure 7 shows $\Pi(p)$ for r=0 and N=1. Here we see how the system does not appear to approach a normal percolation threshold, which looks like a step function for large lattice systems. In figure 7, large system sizes have a soft shoulder for small p. Inspecting individual $\Pi(p)$ profiles, I found that this behavior transitions at r=1.5, thus partially resolving the issue.

Discussion/Conclusion:

I studied site percolation on the Kleinberg small-world model. Using numerical simulations I determined the percolation threshold for a wide range of values of the system's two parameters: the 'clustering coefficient' r and number of long range links N. The percolation threshold values were corrected for finite-size effects. The percolation threshold contour showed several interesting features. First, I found that for large N, N>2, p_c does not change anymore. This result is surprising and warrants further study. I plan to simulate similar systems to see if the result is occurs commonly. Another interesting feature is the limit of small r for N=1 and N=2. In these cases, p_c takes on values of

≈ 0.25 and ≈ 0.20. This might be related to reasoning behind the percolation threshold for the Erdős–Rényi random graph. The values of p_c are close to ¼ and 1/5, which are both $1/(z-1)$, where z is the number of out links.

There are a number of 'special' values of p_c, N, and r. Unfortunately, I have so far been unable to come up with any theories or compelling conjectures for why these values should be important.

Apart from theory, there are some 'practical' things that can be gleaned from the main contour plot. If this model was an accurate description of real human networks, then my results could help make several policy positions in regard to epidemics. First, if people on average tend to have more than two long-distance friends, then it is always better to try to reduce p_c by increasing r. This means convincing people not to contact *very* long distance friends. If the average number of long-distance friends is less than one, then it's always better to try to reduce the average number of long-distance friends. If on the other-hand, the goal is to spread innovation, then the exact opposite set of recommendations would be best.

References:
Kleinberg, J. (2000). "The small-world phenomenon: an algorithm perspective". *Annual ACM Symposium on Theory of Computing*, ppg. 163.

Suding, P. N., Ziff R. M. (1999). "Site percolation thresholds for Archimedean lattices". *Physical Review E* **60** (1): 275–283.

Lorenz, C. D., R. May, Ziff, R. M. (2000). "Similarity of Percolation Thresholds on the HCP and FCC Lattices". *Journal of Statistical Physics* **98**.

Callaway, D.S., Newman, M.E.J., Strogatz, S.H., Watts, D.J., "Network Robustness and Fragility: Percolation on Random Graphs" *Physical Review Letters* **85** 5468.

Newman, M.E.J. Watts, D.J. (1999). "Scaling and percolation in the small-world network model" *Physical Review E* **60**, 7332.

Newman, M.E.J, Jensen, I., Ziff, R.M. (2002) "Percolation and epidemics in a two-dimensional small world". *Physical Review E* **65**, 021904.

Sander, L.M., Warren, C.P., Sokolov, I.M., Simon, C., Koopman, J. (2002) "Percolation on heterogeneous networks as a model for epidemics" *Mathematical Biosciences* **180,** 293.

Figures:

Figure 1: The probability that there exists a spanning cluster of active sites (Π) verses the percentage of sites that are active (p) for various system sizes (W) for r = 4. Note that all of the curves cross at roughly the same point. This is a feature that is seen in normal lattice percolation.
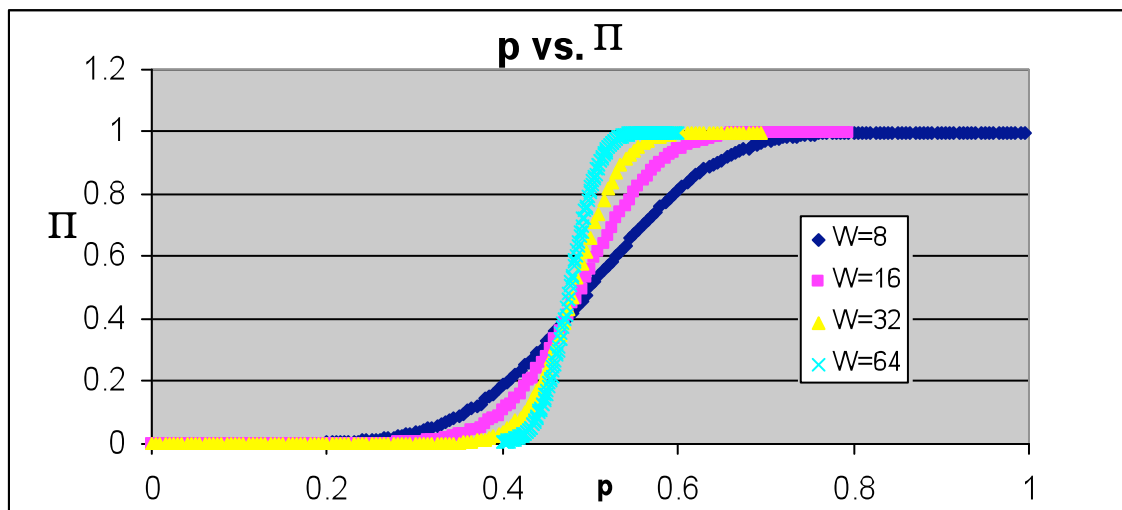

Figure 2: Fitting the convergence to the infinite system size p_c for various system sizes. We see that the relationship is a power-law with an exponent of roughly -0.75, which is the same result seen for finite scaling of normal 2D lattices. The form of the scaling is power-law: $p_c(W) - p_c(\infty) = W^\alpha$.
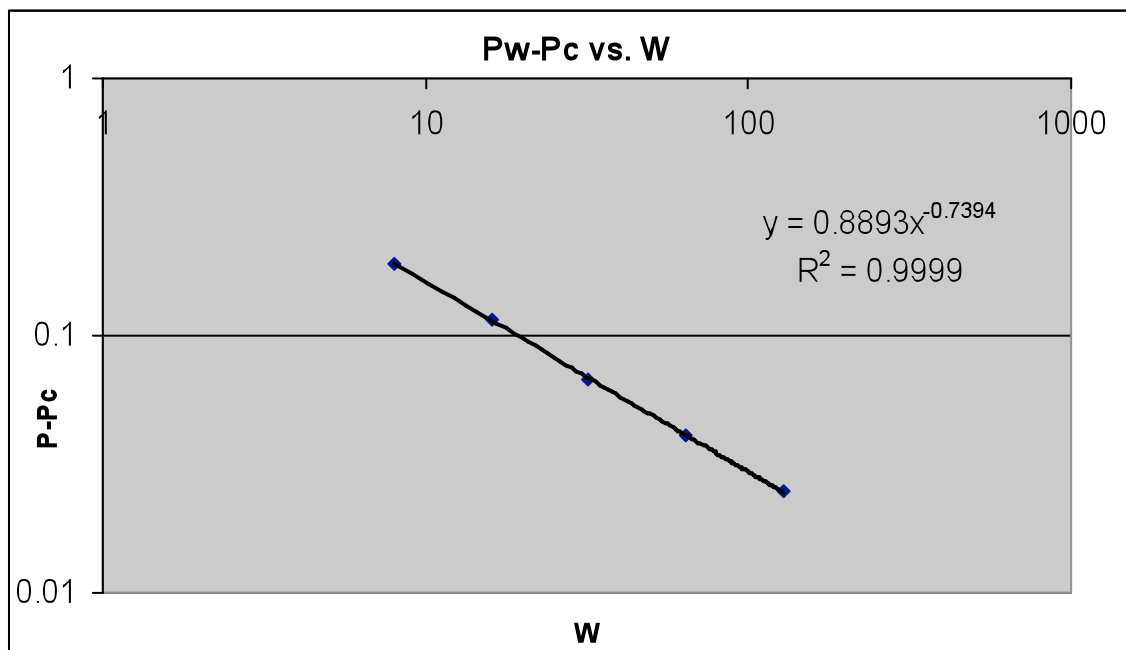
Figure 3: The contour plot of the percolation threshold verses the two parameters of the Kleinberg model, the clustering coefficient r and the number of long range links N.
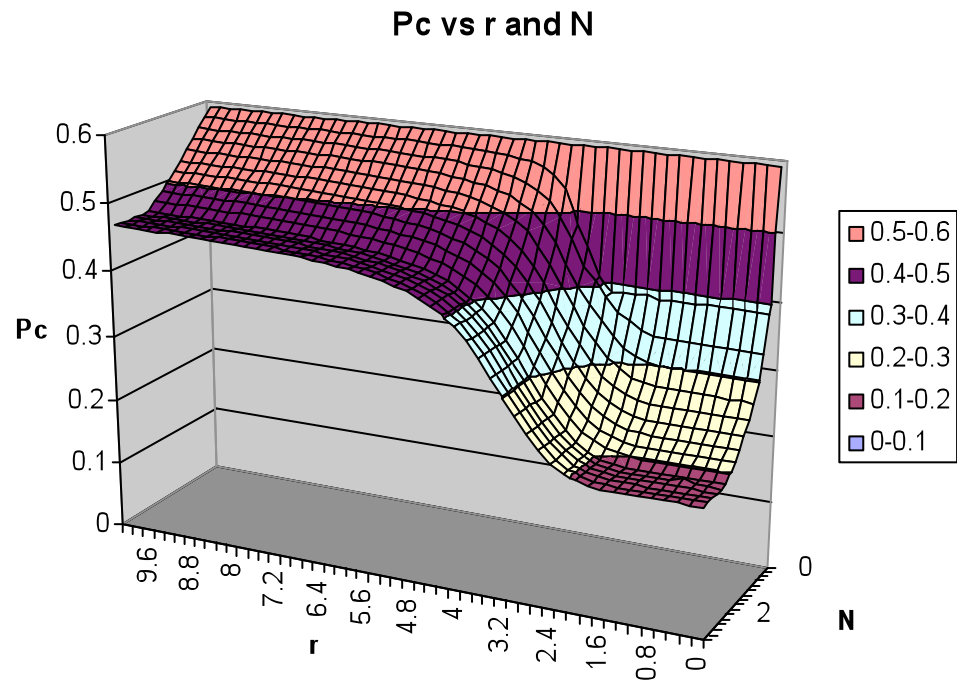

Figure 4: A slice of the contour plot in Figure 3. The slice is at r = 0. The pink function is the results of Newman 2002 for *bond* percolation on a square with shortcuts.
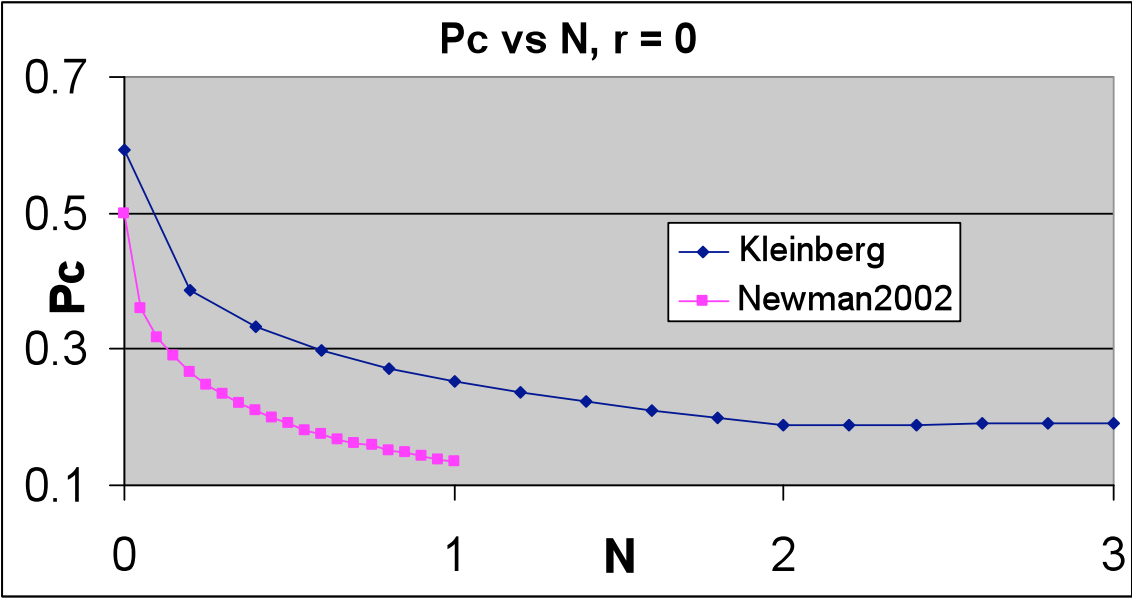
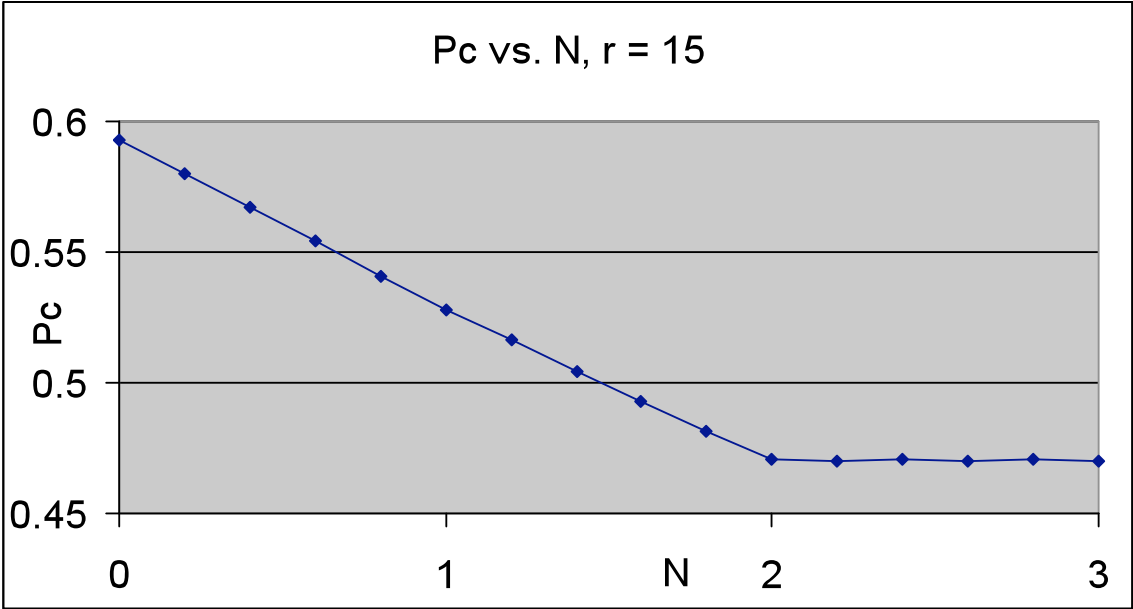Figure 5: A slice of the contour plot in Figure 3. The slice is at r = 15.

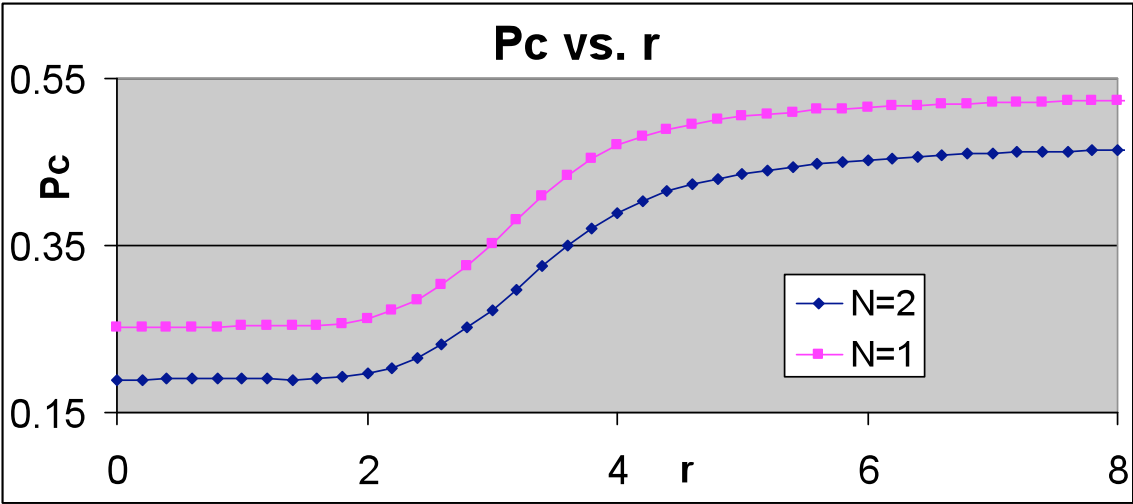Figure 6: Slices of the contour plot in Figure 3. The slices are at N=1 and N=2.

Figure 7: The probability that there exists a spanning group of active sites (Π) verses the percentage of sites that are active (p) for various system sizes (W) for r=0. Note that unlike Figure 1, these curves do not overlap. This happens to be enough to distinguish between the two types of percolation.