



Network Structure and Diffusion

James Moody

June 2009

PWP-DUKE-2009-004

*Duke Population Research Institute
On-line Working Paper Series*

Network Structure and Diffusion*

By

James Moody
Duke University Sociology

June, 2009

Words: 5,567
Tables: 2
Figures: 4

* This work is supported by NIH grants DA12831 and HD41877 and NSF grant SES-0624158. Thanks to the Center for Advanced Studies in the Behavioral Sciences for providing space and time to work on this project. Please direct all correspondence to James Moody at jmoody77@soc.duke.edu

Abstract

The diffusion of information, diseases, or other generic “goods” through a social network is determined by a joint function of the dyadic transmission probability and the overall structure of the network. But what network features matter? While recent theoretical work has highlighted a number of potential features, few have examined diffusion potential in empirical networks. This paper grounds theory on network diffusion by simulating diffusion over observed friendship networks. In addition to the well-known effects of average distance and local clustering, results show that the number of node-independent paths between pairs significantly moderates the effect of average distance.

Introduction

The recent boom in interest about networks is driven in part by the recognition that networks carry things between actors. Hot stock tips quickly move across email networks, rumors of job offers move quickly between colleagues (Granovetter 1973), ideas are exchanged through coauthorship networks (Moody 2004), and diseases such as HIV move through needle exchange or sexual networks (Morris 1997). While networks certainly provide more than a conduit for diffusion (Gould 2002; White, Boorman, and Breiger 1976), it is one of their most direct functions. But if networks matter because they shape diffusion, what aspects of network structure are most important? While theory has proposed a number of network features, there has been relatively little empirical work comparing diffusion rates across real networks. Since real networks contain structural features that are difficult to accurately simulate (Snijders and Van Duijn 1997; Snijders 1990), any understanding of network diffusion based solely on artificial networks will likely be incomplete.

Researchers recognized the importance of network structure for information flow early (Coleman, Katz, and Menzel 1966; Pool and Kochen 1978; Rogerst 1962; Valente 1995; Valente 2001). Coleman's classic study of medical innovation, for example, suggested that physicians with many social contacts were much more likely to use new medicines. Empirical work repeatedly shows that hard-to-find information travels through networks. Lee (1969) showed that finding an abortionist (when it was illegal) rested on the diffusion of information through networks, a finding that shaped Granovetter's famous insight about the strength of weak ties (Granovetter 1973). A large body of research on the diffusion of innovations (Rogerst 1962; Valente 1995; Valente

2001) and peer influence (Friedkin 1998; Friedkin and Cook 1990) has shown that an individual's ideas and behaviors are direct functions of the ideas and behaviors of the people they are connected to.

Work in epidemiology has relied heavily on network diffusion models, particularly for diseases such as HIV, that require close personal contact (see Morris 1993; 1997). Epidemiologists recognized that simple random mixing was inappropriate in these contexts, and built models that accounted for the different likelihood that people with different attributes (like number of sexual partners) would mix (Lajmanovich and Yorke 1976; Hethcote and Yorke 1984; De Gruttola and Mayer 1988) through bias parameters similar to those developed for social networks (Fararo 1981; Fararo and Skvoretz 1987; Pool and Kochen 1978). This work showed that increases in homophily slowed disease spread in simulated networks. Generalizations to the overall degree of assortative (like-with-like) mixing suggest that epidemics spread faster in assortative settings but further in disassortative networks (Morris 1997). Recent work coming out of statistical physics has focused on heterogeneity in number of partners (the degree distribution), suggesting that networks with very skewed degree distributions ("scale-free") will have much larger epidemics.

Substantively, there are two analytically distinct problems in understanding the effect of network structure on diffusion. On the one hand, we can examine variation *within* networks, by asking what positional features increase the likelihood of acquiring whatever is flowing through the network (are central actors more likely to acquire information, for example). As we will see below, diffusion rates *within* empirical networks can vary significantly. Empirically, one can answer this question by simulating

diffusion over an observed network of interest and seeing who is most likely to acquire the good flowing through the network (Rothenberg et al. 1995). On the other hand, we can ask about differences in diffusion *between* networks. The substantive question here is why diffusion progresses faster or moves further in different populations. This problem is challenging empirically, however, as one needs similar data on many networks to gauge diffusion variation. Here, I use data on 125 real-world friendship networks to identify how different network features affect total diffusion levels.

Network Structure and Diffusion

The Diffusion Process

Diffusion over networks is controlled at the *transmission level* by the probability (P_{ij}) that a person with the “good” passes it to contacts that do not have it at each time point.¹ Empirically, the specific value of P_{ij} likely varies substantially by the type of good and the type of node. To ensure that differences across networks are due only to changes in the structure of the network, in the simulations reported below I hold P_{ij} constant over all dyads. As time passes, those with the good pass it to their contacts, who have the opportunity to pass to their contacts in the next round.

Given P_{ij} , one can think intuitively about the probability of diffusion over the entire network as a joint probability problem linking dyads together in long chains. For

¹ For simplicity, I will refer to whatever is moving through the network as a “good.” Of course, many times the “good” is decidedly bad, as when it is an infectious disease. While the social interaction processes for the transmission of different types of goods likely varies systematically by the type of good, the resulting variation can in most cases be reduced to variance in P_{ij} . For the purposes of this paper, where the goal is to identify the independent role of network topology on differences in diffusion across networks, I ignore the channeling effect of relationship timing (Morris and Kretzschmar 1997, Moody 2000). In practice, relational timing can have a dramatic effect on diffusion, as anything less than complete concurrency will always generate lower diffusion levels than observed here (Moody 2000). The various interactions of timing and structure in shaping diffusion are examined in detail elsewhere (Moody 2006).

example, the probability that i passes the good to j who then passes to k would be $P_{ij}P_{jk}$. Since P_{ij} is always less than 1, the likelihood of diffusion drops as the number of dyads in the chain increases. On the other hand, the probability increases as the number of potential routes increases, so if i could also pass to k through m , then the probability of the good reaching j is the probability that it goes through one path or the other (minus the probability that it goes through both). *That is, the probability decreases with distance and increases with the number of alternative routes.* While this probability calculation is tractable for small numbers of independent chains, these products become intractable as the complexity of the network increases.

At the network level, theorists have identified 3 categories of features thought to shape diffusion over static networks. *Connectivity* refers to the system of paths formed by the concatenation of local networks. These paths are the long-distance transmission lines that carry a good from an initial node to others and their arrangement determines global diffusion. *Clustering* refers to the likelihood that a path starting from one node returns back to the starting node, and variants refer to just how quickly paths fold in on themselves. *Degree features* of the network refer to the amount and pattern of direct contact, such as the distribution of number of contacts and the extent of assortative mixing.²

² Theoretically, degree features should shape clustering and connectivity. But there are three reasons not to subsume degree features under these headings. First, I think it's an empirical question whether degree features have any *independent* effect on diffusion net of the connectivity and clustering features. That is, while assortative mixing shapes things like clustering and distance and skewed degree distributions create short paths, they might do so in a manner that magnifies the path patterns. Second, degree features are remarkably simple to measure with only local data. So if we can find a systematic effect, this will be of practical value. Finally, much of the recent theoretical and simulation work on networks has focused on degree features, and we want to be consonant with this work.

Connectivity

A path in a network is any sequence of connected nodes and edges that starts and ends with a different node ($i \rightarrow j \rightarrow k \rightarrow l$). A *component* of a network is a maximal set of nodes that has at least one path connecting every pair of actors in the component. If transmission were certain, then every member of a component would get any good introduced into the component. As such, components form the maximum diffusion potential in a network, as diffusion cannot reach people in different components. This feature is fundamental for very sparse networks, such as sexual networks, where the average number of ties is quite small. In such cases, it is likely that any given population will be composed of many small components. However, there is a very sharp threshold based on number of ties, where the network component structure cascades such that the vast majority of people are in the largest “giant” component (Molloy and Reed 1998; Newman et al. 2001; Palmer 1985). In graphs with relatively high degree levels (such as the networks examined here) between 90% and 95% of all nodes will be in the giant component.

Within components, the *path distance* between any pair of nodes is defined as geodesic distance: the number of edges on the shortest path connecting the pair. Thus two friends are at distance 1, while friends of friends are at distance 2 and so forth. The path length (L) for a given network is the average geodesic distance between all pairs of actors in the network. In a random network, L scales with the log of the number of nodes (Newman 2001). As demonstrated by Watts (Watts 1999; Watts and Strogatz 1998), it only takes a small number of random connections in a network to bring observed path lengths in the range of random graphs with similar contact volume. Because these

random ties form shortcuts that lower distance over the entire network, *we expect diffusion rates to decrease with distance.*

While L captures the decrease in transmission probability through compounding probability chains, the *number of node-independent paths* captures the additive effect of alternate routes in the network (Moody and White 2003). In any connected component, there exists a set of nodes S that, if removed, would disconnect the graph. For any good to pass from one side of this cutset to the other, the good must travel through the nodes in S . Two paths are node-independent if they only have their starting and ending nodes in common. Menger proved that the number of node-independent paths connecting any pair of actors in the network is equal to the size of the smallest cutset in the graph, a duality that defines structural cohesion in a network (Moody & White 2003). The number of node-independent paths also determines the set of (likely nested) k -components. By definition, any simple (1-)component has at least 1 path connecting every pair of actors in the network. A bicomponent has at least 2 node-independent paths, a tri-component has 3 node independent paths and so forth. Because each independent path represents a completely unique route for diffusion, in general *we expect that diffusion rates increase with the average number of node-independent paths connecting pairs in the network.*

The effects of distance and number of paths should not be independent, however, since every path has to have a distance. If you have many short paths (2 to 3 steps, say) between pairs of nodes you substantively have a close-knit cluster. As such, diffusion should move quickly across these multiple short paths to spread diffusion *within* the cluster, but diffusion will not spread quickly beyond this group. On the other hand, if you have multiple long distance paths, then you should get greater diffusion over the

entire network. As such, we expect an interaction between average distance and number of paths, such that *the diffusion-reducing effects of distance decrease as the number of node-independent paths increases.*

Clustering

The second basic feature thought to shape diffusion in networks rests ultimately on the *efficiency* of the path structure. It is useful to think of diffusion as a tracing procedure through a network. Starting from a single “infected” node, one then traces out to this node’s direct contacts, then from them to their contacts and so forth. The most efficient (though not necessarily robust) diffusion structure is a network tree: where each new tie in the network connects to a new node. Efficiency is reduced when network paths loop back on themselves, as when a friend of a friend is a friend. These closed loops mean that many of the edges in the network lead “back” to nodes already reached by the infection. This process slows diffusion, as the good is effectively trapped in dense local clusters.

The most direct way to measure this kind of redundancy is with the amount of *local clustering* in the graph, defined as the proportion of completely connected triads divided by the number of triads with 2 or 3 edges. Early researchers (Holland and Leinhardt 1970; Holland and Leinhardt 1971; Pool and Kochen 1978) called this the transitivity level, more recent work has called it the “clustering coefficient,” (Newman 2000; Watts 1999). This term is somewhat unfortunate since while all intuitively “clustered” networks have high transitivity levels, not all graphs with high transitivity levels are intuitively clustered (think of long lattice-like sequences, for example). The

clustering ratio (C) captures the purely local redundancy in the network. *As C increases, we expect diffusion rates to decrease.*

While C captures local redundancy, networks often form much larger community structures, where the probability of a tie falling within some group is much higher than the likelihood that a tie would fall between groups. For example, we know that friendships among adolescents are much more likely to fall within grades than between grades. These foci (Feld 1981) will generate closed loops that are larger than triads. While these community factors are often based on attributes (race or grade, for example), they can also emerge endogenously and appear as primary social groups. Because these collections of nodes channel paths back to prior nodes, *we expect that diffusion rates decrease as community-level clustering increases.*

Degree Features

Past theory and research on network diffusion has focused on two (related) elements of the number of direct contacts. Early work in network epidemiology highlighted the extent to which people with similar numbers of partners had ties to each other. The degree of *assortative mixing* reflects homophily based on number of ties. For example, in sex networks ties between commercial sex workers and their clients should create very disassortative mixing. Such patterns generate star-like networks that provide an efficient diffusion setting. The extent of assortative mixing can be measured simply as the degree correlation across on all dyads.

Recent work on large-scale networks has suggested that many (perhaps most) social networks have highly skewed degree distributions (Barabasi and Albert 1999; Dezso and Barabasi 2001). These networks have degree distributions where most nodes

have few contacts and a small number have many orders of magnitude more. When the degree distribution is plotted on a log-log scale, the distribution forms a straight line, and is often said to have a “power-law” distribution, since $p(k) \sim k^{-\lambda}$, where λ is the power exponent (see Jones and Handcock 2003 for a critique of this work). If the network is constructed through a preferential attachment process, where the probability of a newly entering node connecting to a current node is proportional to the degree of the current node, then scale-free networks will have star-like structures, and thus short average distances. *As such, we expect that diffusion rates increase as the degree distribution becomes more skewed.*

The preferential attachment model is reminiscent of early epidemiology work that focuses on degree-based mixing, but focuses directly on the mixing mechanism (instead of through the resulting degree distribution). The short-distance, star-like networks are more likely when we have highly disassortative mixing: where a small number of stars have many contacts, all of which are only connected to the star. In these networks, diffusion will likely spread far, but somewhat slowly, hopping from star to star. On the other hand, the network can be composed of assortative mixing (equivalent to the proportional mixing models) where people with many ties are tied together. Here, diffusion will spread quickly. *Thus, as the degree correlation increases we expect faster diffusion rates.*

Data and Methods

To evaluate these various mechanisms described above, I simulate diffusion across 125 observed friendship networks. I then regress the diffusion rates (detailed

below) on the network structure measures. Next I describe the network sample, the diffusion simulation and the resulting diffusion curve-based measures.

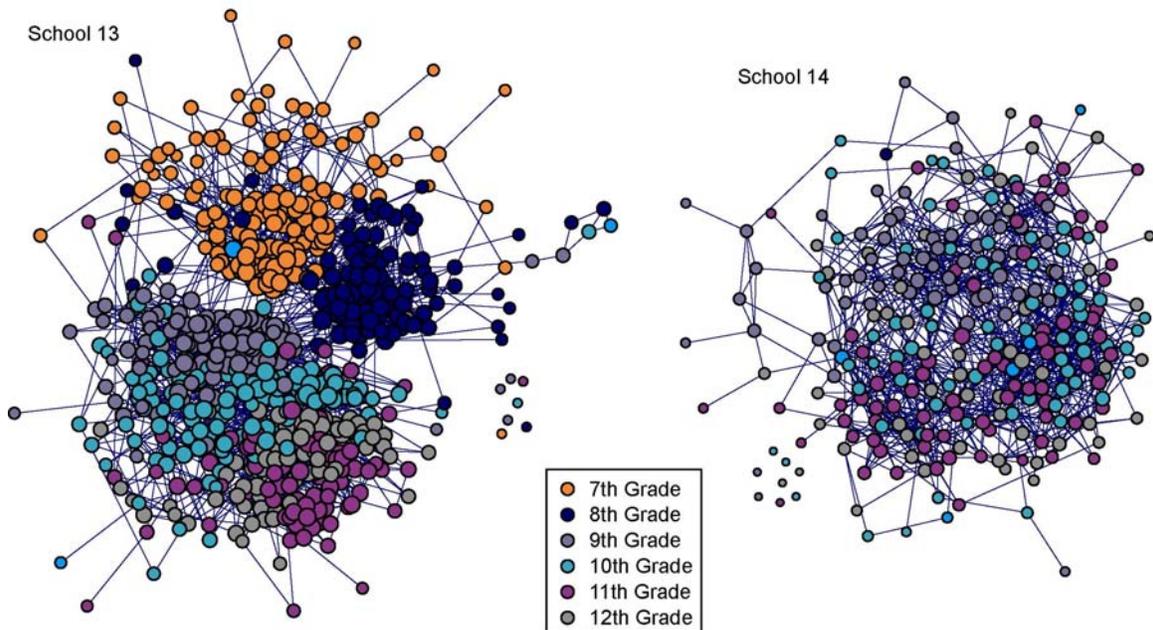
The Add Health Networks

Data for this project come from the in-school friendship networks collected by the National Longitudinal Survey of Adolescent Health (Add Health) in 1994. Add Health surveyed adolescents in 80 communities, sampling a single high school (with a 10th grade) and (if separate) a Jr. high “feeder” school, resulting in an ultimate sample of 90,118 adolescents. Due to sampling and roster-coverage issues in the original sample, 15 schools (comprising 7489 individuals) with low sampling rates (less than 50%) were excluded from the sample. An additional 6758 cases cannot be used because the adolescents did not have a nominatable ID (typically these were students who moved into the district after the rosters were printed), meaning that they could not be named as friends. Combined, this results in a “network sample” of 129 schools and 75,871 adolescents. Of these schools, I further remove 1 school because it has only one grade (making it impossible to calculate same-grade homophily measures), one because the number of out-of-school nominations was inordinately high, creating many small components, and two others that were very small (less than 30 nodes), for a resulting sample of 125 networks.

Adolescents were asked to name their 5 best male and 5 best female friends, and could nominate friends outside of the school (though these nominations could not be matched). For the purposes of this paper, I treat all nominations as undirected, which allows for a much greater degree variance. The advantage of this dataset for studying diffusion processes is that collecting networks with a similar instrument in 125 different

settings allows a good deal of real-world variation in network structure, making it possible to identify those features that affect rates of spread across networks. Figure 1 below presents example networks from two schools.

Figure 1. Example Sociogram
Node size is proportional to $\ln(\text{degree}+1)$.

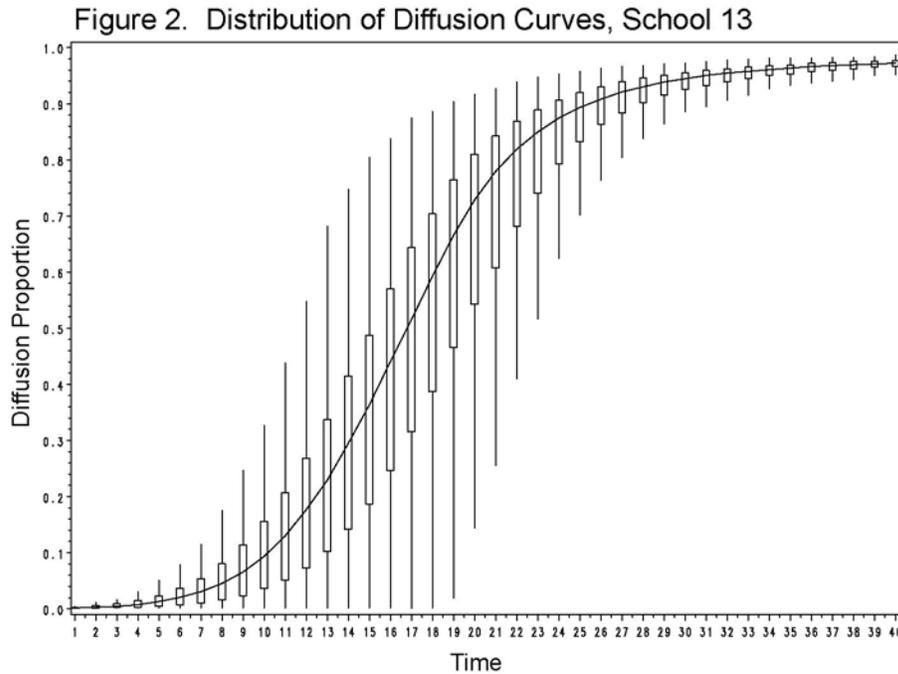


Both of these are schools in the Midwest and majority white. School 13 has 685 nodes and spans both high school and Jr. high school. School 14 is smaller, with just 268 nodes. Colors denote grade in school, and it is evident that cross-grade mixing is much higher in school 14 than in school 13.

Measuring Diffusion

I simulate diffusion over each school network to measure diffusion potential. Each simulation starts by randomly selecting a single start node from the network. At each iteration, those with the good pass the good on to their contacts without the good with probability P_{ij} , here held constant at 0.08. At each iteration, the new proportion of

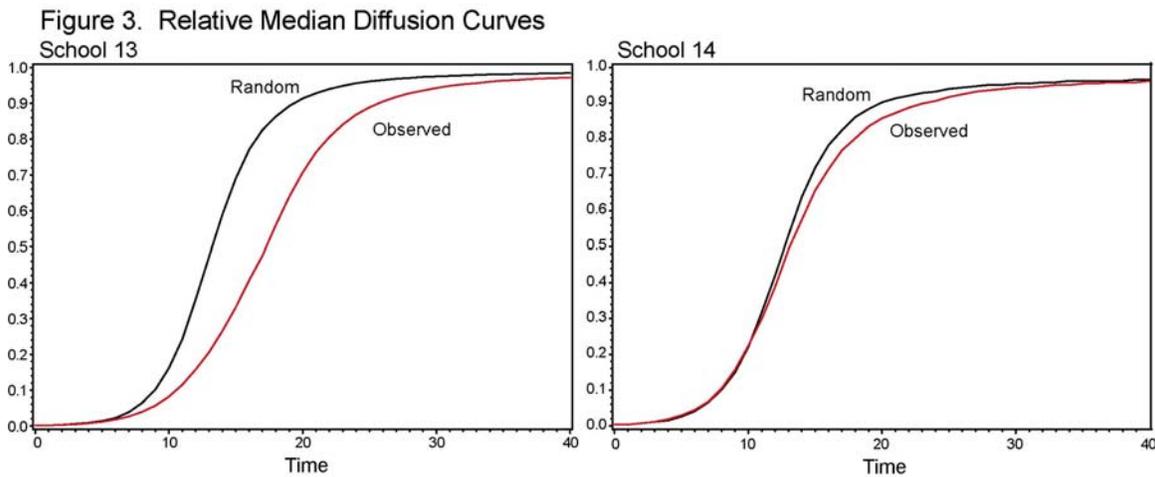
the population who have the good is recorded, generating a cumulative diffusion curve. This process is repeated 500 times, generating a distribution of diffusion curves for each network. I use the median proportion with the good to define the diffusion curve rate in each network. Figure 2 below shows the distribution of observed diffusion curves for school 13.



The solid line represents the median proportion with the good, while the boxes capture the inner-quartile range and the whiskers stretch to any data points within 1.5 times the IQR. As is clear from this figure, there is a good deal of internal variation in the diffusion levels, which results from variance in the position of the starting node and random chance.

To effectively compare diffusion rates across networks, we need a baseline that lets us distinguish the effect of structure. The obvious baseline is to compare the observed diffusion curve to that found in a similar structureless network. I generate random networks with matching degree distributions (and thus density) and repeat the

diffusion simulation over these random nets (I use the network generating algorithm described in Newman et al. 2001). For each observed network, I generate 10 matching random networks and simulate the diffusion process 50 times in each network, for a matching 500 observations. This random matching procedure ensures that any differences in the diffusion process are due solely to the structural features of the observed network. In addition, the random baseline comparison minimizes any potential difficulties associated with choosing P_{ij} , since it is held constant in both cases.³ Figure 3 below compares the observed median diffusion rates (red line) to the random diffusion rates (black line) in both of the example networks given in figure 1.



While the distance varies, in all cases the observed curve fall under the random curve, indicating that diffusion reaches more people faster in random networks than in the observed networks, which is exactly what we would expect from the literature on biased networks. Since real-world networks have community structure, paths are re-routed in on themselves creating redundancy. To capture the relative difference between networks, I compare the random diffusion curve to the observed curve using the ratio of the area

³ I've run the simulations using $P_{ij} = 0.20$ as well. The results are essentially identical, which is what we would expect given the high correlation (0.97) between the relative curve measure for the two P_{ij} levels.

under the observed curve to the area under the random curve. As this measure approaches 1, diffusion spreads as rapidly in the observed network as in the random network. For the two networks given in the example figures above, school 13 has a diffusion curve ratio (**DCR**) of 0.85 and school 14 has at DCR of 0.97.

Independent Variables

I use 2 measures for *connectivity*: the average path length (L) and the average number of node-independent paths between every pair of actors in the network.⁴ Average path length is simply the mean geodesic distance between connected pairs in the network. The number of node-independent paths was calculated using a pair-level variant of the algorithm presented in Moody & White (2003).

Clustering properties are measured locally with the clustering coefficient (transitivity ratio), which is the proportion of all connected triples that are closed triads, calculated directly from the triad census (Moody 1998; Wasserman and Faust 1994). I use two measures for clustering beyond this local level. Because grade in school is among the most dominant foci in schools, I use the log-odds of a nomination falling within grade in each school. Second, I measure the relational strength of local peer groups. I first identified primary subgroups in each network using an algorithm that simultaneously maximizes the ratio of in-group to out-group ties while ensuring that the group is at least biconnected (Moody 1999; Moody & White 2003). Once such groups are identified, I use the mean segregation index (Freeman 1972) for each group. If all of a group's ties fell within the group, the value would be 1.0. If ties were distributed at random, the value would be 0. Thus schools with high values have local peer-groups

⁴ Computationally, both means are taken from a random sample to save time.

that are strongly segregated. Substantively, this measure tends to capture the additional community clustering that occurs *within* grades.

Two degree-based measures capture heterogeneity in network involvement and assortative mixing. I use the skew of the degree distribution to measure the potential effect of a long-tailed degree distribution. To measure homophily on number of ties, I use the dyad-level correlation of the degree of node i with degree of node j .

Finally, in all models I control for network size (needed to scale distance and path measures), the proportion of nodes in the network that are isolated, and flag any network where the maximum diffusion curves did not reach more than 90% of the population (these are, typically, networks with many isolates).

Table 1 gives the description and descriptive statistics for all of the measures used in this paper.

Table 1. Descriptive Statistics

Name	Description	Mean	StD	Min	Max
Dependent					
Diffusion Curve Ratio	Area under the observed median diffusion curve divided by area under the degree-matched random network median diffusion curve.	0.89	0.08	0.55	1.02
Independent					
<i>Connectivity</i>					
Distance	Average distance between connected pairs.	3.87	0.71	1.79	5.49
Independent Paths	Average number of node independent paths connection each pair in the network.	4.88	1.13	2.32	7.17
<i>Clustering</i>					
Clustering Coefficient	Proportion of connected sets of 3 that are closed triads (=T300/(T300+T201))	0.08	0.04	0.03	0.32
Grade Homophily	Log of the odds that a nomination falls within grade.	2.30	0.52	0.96	4.21
Peer Group Strength	Average segregation index for identified sociometric peer groups.	0.93	0.04	0.69	0.98
<i>Degree Distribution</i>					
Degree Skew	Skew of the degree distribution	0.65	0.28	-0.25	1.27
Assortative Mixing	Degree correlation across dyads	0.21	0.06	0.00	0.39
<i>Control Variables</i>					
Network Size	Number of nodes in the network	595	439	25	2250
Proportion Isolated	Proportion of nodes with no ties	0.03	0.04	0.00	0.23
Non-Complete	Dummy: = 1 if the maximum median diffusion curve is < 0.80	0.07	0.26	0	1
N Networks:	125				

Results

I start by assessing the effect of each domain separately and then model them simultaneously. Table 2 presents these results.

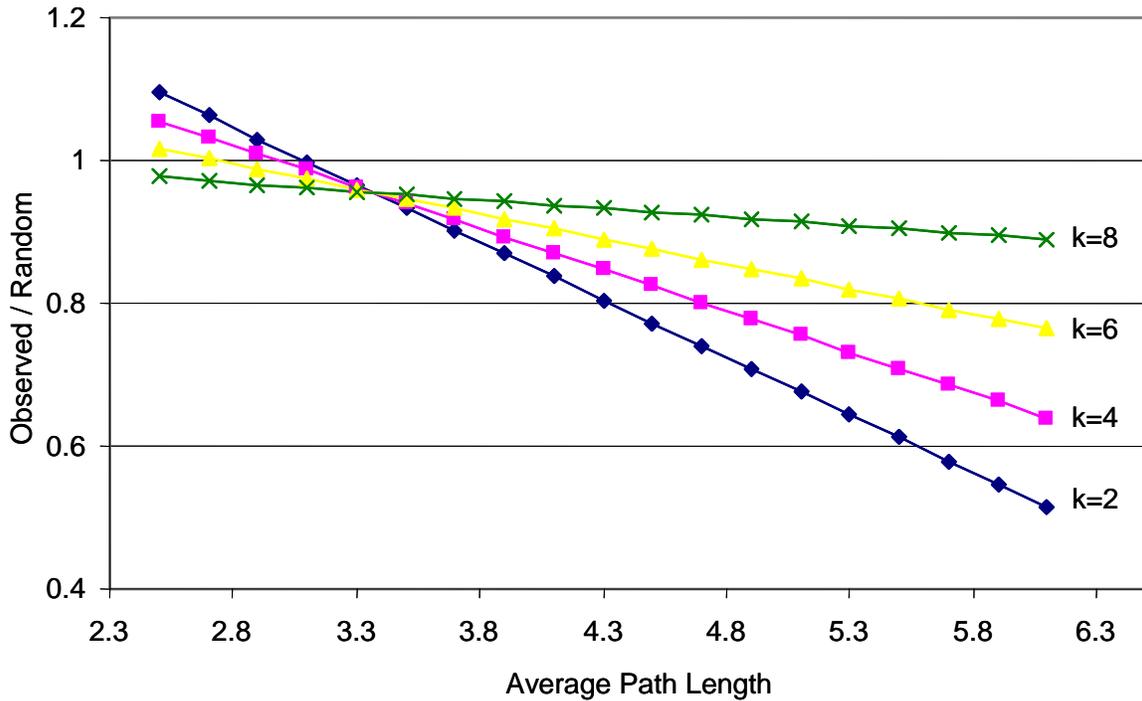
Table 2. OLS Regression of Relative Diffusion Ratio on Network Structure

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	1.62***	1.90***	1.02***	1.81***	1.71***
<i>Connectivity</i>					
Distance	-0.207***			-0.179***	-0.171***
Independent Paths	-0.077***			-0.056***	-0.052***
Distance x Paths	0.023***			0.015***	0.016***
<i>Clustering</i>					
Clustering Coefficient		-0.692***		-0.653***	-0.454***
Grade Homophily		-0.026**		-0.007	-0.009*
Peer Group Strength		-0.868***		-0.141	-0.146
<i>Degree Distribution</i>					
Degree Skew			-0.023	-0.007	-0.002
Assortative Mixing			-0.189*	-0.059	-0.071
<i>Control Variables</i>					
Network Size/100	0.005***	-0.005***	-0.005***	0.004*	0.002**
Proportion Isolated	-0.007	-1.106***	-0.984***	-0.300**	0.058
Non-Complete	-0.006	-0.052*	-0.078**	-0.006	0.018
Adj- R ²	0.85	0.76	0.60	0.90	0.93
N	125	125	125	125	121

Model 1 shows the independent effects of the connectivity features. Here we see that as average distance increases, the relative diffusion ratio drops, as predicted.

Moreover, the interaction effect with the number of unique paths is strong and clear. To make this effect clear, the interaction is plotted in figure 4 below. The strong negative effect of distance is attenuated as the number of independent paths increases.

Figure 4. Relative Diffusion Ratio
By Distance and Number of Independent Paths



Model 2 shows the direct effect of the observed clustering features. As expected, all of the clustering features have a strong negative effect on diffusion rates. As judged by the standardized coefficients, the effect of grade segregation is the weakest of the 3 measures (a 1 standard deviation increase in the log-odds of nominations within grade decreases the relative diffusion ratio by .16 standard deviations), and the effect of local clustering (standardized beta = -0.34) and peer-group clustering (standardized beta = -0.48) are significantly stronger. In general, as the network becomes more clustered, diffusion across the entire network slows.

Model 3 shows the direct effects of the degree distribution features. The degree skew has no appreciable effect on the diffusion rates. This is probably not surprising given that the random networks were generated with a matching degree distribution, so any effect of the distribution itself has been captured in the random network. The

assortative mixing parameter has a significant effect, and networks that are more assortative with respect to degree have lower overall relative diffusion ratios.

Model 4 shows the net effect of each measure, and substantively reflects the fact that the primary graph-theoretic factors of clustering, distance and number of independent paths largely determine the diffusion rates. The effect of grade-based clustering and peer-group clustering is no longer significant, which is likely due to the fact that such foci shape the distribution of paths and distance in the network. Examination of the residuals plots suggested that 4 networks were significant outliers. Model 5 shows the results with these networks removed. The general story is the same, but the effect of grade clustering becomes a little stronger and the effect of the clustering coefficient a little weaker.

Summary & Conclusion

These simulations suggest that path distance mediated by the number of paths and local clustering are crucial for understanding between-network variance in relative diffusion risk. While much recent work has focused on the effects of network shortcuts in large-scale, highly clustered networks (Watts 1997), comparatively little work has focused on how alternate routes make networks more robust for diffusion (Moody & White 2003). Here we see that as the average number of independent paths between nodes increases, diffusion is more likely to occur over long paths. This will be particularly relevant in contexts, such as those recently hypothesized for belief in urban legends (Centola, Eguiluz, and Macy 2006), where adoption requires receiving the good from multiple contacts independently.

This study focused on features affecting overall diffusion across different populations. I would expect the positional correlate of multiple connectivity to have a

similar robustness effect within networks. That is, if an “infected” and “receptive” node were jointly connected by multiple paths, the distance separating them should become less relevant. Moving to the local level is complicated, however. In social settings such as those examined here, we would expect that nodes deeply nested within clusters will be at high risk to diffusion within clusters, but relatively protected between clusters. How this result interacts with purely local clustering will be the focus of future work on within-network diffusion variability.

Finally, the results here suggest no effect of the skew of the degree distribution. Recent work on scale-free networks would suggest that highly skewed networks would have particularly fast diffusion rates. We have to interpret this result carefully, however, since the networks examined here have low variability in degree skewness and the random comparison graphs are similarly skewed. These factors will likely minimize any effect of degree skew.

Examining the effect of graph structure on diffusion potential in ideal-type graphs simulated to fit particular models (such as small-world networks or scale-free graphs), allows us to identify the way diffusion works in idealized settings. Most real networks, however, are often complex meldings of multiple models, where idiosyncratic features of simulated networks may not be well matched. This paper used the natural variation in real networks as a baseline for comparing the effects on diffusion potential of multiple hypothesized factors, showing clear evidence for the effects of distance mediated by number of independent paths and local clustering. Future work will extend this project to examine within-network diffusion and the robustness of diffusion when the graph is dynamic.

Reference List

- Barabasi, Albert-Laszlo and Reka Albert. 1999. "Emergence of Scaling in Random Networks." *Science* 286:509-12.
- Centola, D., V. Eguiluz, and M. Macy. 2006. "Cascade Dynamics of Multiplex Propagation." Forthcoming, *Physica A: Statistical and Theoretical Physics*.
- Coleman, James S., E. Katz, and H. Menzel. 1966. "The Diffusion of an Innovation Among Physicians." *Sociometry* 20:253-70.
- Dezso, Zoltan and Albert-Laszlo Barabasi. 2001. "Can We Stop the AIDS Epidemic?" cond-mat/0107420. <http://xxx.lanl.gov/archive/cond-mat>.
- Fararo, Thomas J. 1981. "Biased Networks and Social Structure Theorems." *Social Networks* 3:137-59.
- Fararo, Thomas J. and John Skvoretz. 1987. "Unification Research Programs: Integrating Two Structural Theories." *American Journal of Sociology* 92:1183-209.
- Feld, Scott L. 1981. "The Focused Organization of Social Ties." *American Journal of Sociology* 86:1015-35.
- Freeman, Linton. C. 1972. "Segregation in Social Networks." *Sociological Methods and Research* 6:411-30.
- Friedkin, Noah E. 1998. *A Structural Theory of Social Influence*. Cambridge: Cambridge.
- Friedkin, Noah E. and Karen S. Cook. 1990. "Peer Group Influence." *Sociological Methods and Research* 19(1):122-43.
- Gould, Rodger. 2002. "The Origins of Status Hierarchies: A Formal Theory and Empirical Test." *American Journal of Sociology* 107:1143-78.
- Granovetter, Mark. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 81:1287-303.
- Holland, Paul. W. and Samuel Leinhardt. 1970. "A Method for Detecting Structure in Sociometric Data." *American Journal of Sociology* 70:492-513.
- . 1971. "Transitivity in Structural Models of Small Groups." *Comparative Groups Studies* 2:107-24.
- Jones, James and Mark Handcock. 2003. "Sexual Contacts and Epidemic Thresholds." *Nature* 423:605-6.
- Lee, Nancy H. 1969. *The Search for an Abortinist*. Chicago: University of Chicago Press.
- Molloy, Michael and Bruce Reed. 1998. "The Size of the Largest Component of a Random Graph on a Fixed Degree Sequence." *Combinatorics, Probability and Computing* 7:295-306.
- Moody, James. 1998. "Matrix Methods for Calculating the Triad Census." *Social Networks* 20:291-99.

- . 2000. "Indirect Connectivity and STD Infection Risk: The Importance of Relationship Timing for STD Diffusion." *Social Forces* 81:25-56.
- . 2004. "The Structure of a Social Science Collaboration Network: Disciplinary Cohesion From 1963-1999." *American Sociological Review*. 69:213-238
- . 2006. "Watching Time Fly: Diffusion and Visualization of Dynamic Networks" American Sociological Association presentation, Montreal Canada, August.
- Moody, James and Douglas R. White. 2003. "Social Cohesion and Embeddedness: A Hierarchical Conception of Social Groups." *American Sociological Review* 68:103-27.
- Morris, Martina. 1993. "Epidemiology and Social Networks: Modeling Structured Diffusion." *Sociological Methods and Research* 22:99-126.
- . 1997. "Sexual Networks and HIV." *AIDS 97: Year in Review* 11(Suppl A):S209-S216.
- Morris, Martina and M. Kretzschmar. 1997. "Concurrent Partnerships and the Spread of HIV." *AIDS* 11:641-48.
- Newman, M. E. J. 2000. *Models of the Small World*. Sante Fe Institute Technical Paper.
- Newman, M. E. J., S. J. Strogatz, and D. J. Watts. 2001. "Random Graphs With Arbitrary Degree Distributions and Their Applications." *Phys. Rev. E* .
- Newman, Mark E. J. 2001. "The Structure and Function of Networks." .
- Palmer, Edger N. 1985. *Graphical Evolution: An Introduction to the Theory of Random Graphs*. New York: John Wiley and Sons.
- Pool, Ithiel d. S. and Manfred Kochen. 1978. "Contacts and Influence." *Social Networks* 1:5-51.
- Rogerst, E. M. 1962. *Diffusion of Innovations*. New York: Free Press.
- Rothenberg, R. B., J. J. Potterat, W. W. Woodhouse, S. Q. Darrow, S. Q. Muth, and A. S. Klov Dahl. 1995. "Choosing a Centrality Measure: Epidemiologic Correlates in the Colorado Springs Study of Social Networks." *Social Networks: Special Edition on Social Networks and Infectious Disease: HIV/AIDS* 17:273-97.
- Snijders, T. A. B. and M. A. J. Van Duijn. 1997. "Simulation for Statistical Inference in Dynamic Network Models." Pp. 493-512 in *Simulating Social Phenomena*, R. Conte, R. Hegselmann, and P. Terna. New York: Springer.
- Snijders, Tom A. B. 1990. "Testing for Change in a Digraph at Two Time Points." *Social Networks* 12:163-74.
- Valente, Thomas W. 1995. *Network Models of the Diffusion of Innovations*. Cresskill, NJ: Hampton Press.
- Valente, Thomas W. 2001. "Models and Methods for Studying the Diffusion of Innovations." *Recent Advances in Network Analysis* , P Carrington, Wasserman S., and J. Scott.

Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis*. Cambridge: Cambridge University Press.

Watts, Duncan J. 1999. "Networks, Dynamics, and the Small-World Phenomenon." *American Journal of Sociology* 105:493-527.

Watts, Duncan J. and Steven H. Strogatz. 1998. "Collective Dynamics of 'Small-World' Networks." *Nature* 393:440-442.

White, Harrison C., Scott A. Boorman, and Ronald L. Breiger. 1976. "Social Structure From Multiple Networks I." *American Journal of Sociology* 81:730-780.