The Importance of Relationship Timing for Diffusion*

JAMES MOODY, *The Ohio State University*

Abstract

Relationship timing can have dramatic effects on diffusion through a network, as relationship order determines transmission routes. Though past research has modeled diffusion through static networks or developed methods for modeling change in network patterns, none has combined these factors to show how relationship change channels diffusion. This article formalizes the diffusion problem in networks with changing relations, identifies minimal bounds needed to measure diffusion potential in such networks, and provides a method for identifying who is at risk for diffusion. The effect of timing for diffusion potential is demonstrated with potential flow of sexually transmitted diseases (STDs) in an adolescent romantic network.

* *This research uses data from* the Add Health Project, *a program project designed by J. Richard Udry (PI) and Peter Bearman and funded by grant P01-HD31921 from the National Institute of Child Health and Human Development to the Carolina Population Center, University of North Carolina at Chapel Hill, with cooperative funding participation by the National Cancer Institute; the National Institute of Alcohol Abuse and Alcoholism; the National Institute on Deafness and Other Communication Disorders; the National Institute on Drug Abuse; the National Institute of General Medical Sciences; the National Institute of Mental Health; the National Institute of Nursing Research; the Office of AIDS Research, NIH; the Office of Behavior and Social Science Research, NIH; the Office of the Director, NIH; the Office of Research on Women's Health, NIH; the Office of Population Affairs, DHHS; the National Center for Health Statistics, Centers for Disease Control and Prevention, DHHS; the Office of Minority Health, Centers for Disease Control and Prevention, DHHS; the Office of Minority Health, Office of Public Health and Science, DHHS; the Office of the Assistant Secretary for Planning and Evaluation, DHHS; and the National Science Foundation. Persons interested in obtaining data files from the National Longitudinal Study of Adolescent Health should contact Add Health, Carolina Population Center, 123 West Franklin Street, Chapel Hill, NC 27516-3997 (e-mail: addhealth@unc.edu). Special thanks to Peter Bearman, Martina Morris, Pip Pattison and Lisa Keister, Susanne Bunn, Jill Burkart, and anonymous reviewers for comments on earlier drafts of this article. Previous versions of this article were presented at the IUSSP conference on Partnership Networks and the Spread of HIV and Other Infections, Chiang Mai, Thailand, the Centers for Disease Control in Atlanta, and the American Sociological Association meeting in Anaheim. I thank the conference organizers and participants for helpful comments and suggestions. Direct correspondence to James Moody, Department of Sociology, 300 Bricker Hall, 190 North Oval Mall, Ohio State University, Columbus, OH 43210-1353.*

Social Forces, September 2002, 81(1):25-56

Networks matter, at least in part, because resources and risks flow through relations. People pass rumors, information, drugs, diseases, support, good cheer, or bad moods when they interact with others, who later pass these resources (or risks) to others. Past research on diffusion through social relations shows that a network's global structure guides diffusion. Topological features, such as centrality and clustering, largely shape the diffusion of goods[1] through a network (Burt 1987; Coleman, Katz & Menzel 1966; Kerckhoff & Back 1968; Morris 1993; Myers 2000; Valente 1995). We know, for example, that people whose contacts are not connected to each other are more likely to be exposed to a wider variety of goods than are those with highly clustered networks and that having a few sparse ties in an otherwise clustered network will facilitate diffusion (Watts 1999a). These findings are as applicable to information seekers on Wall Street as they are to nurses who bridge patients with different diseases or small-town gossips seeking tidbits for the next bridge game. In almost every case, however, the networks considered in this line of research are static.

Recent work, however, has shown that social networks, across a wide range of substantive settings, are not static (Carley 1991; Doreian et al. 1996; Doreian & Stockman 1996; Hallinan & Williams 1987; Leenders 1996; Morris & Kretzschmar 1995; Sanil, Banks & Carley 1995; Smith & White 1992; Suitor, Wellman & Morgan 1997; Weesie & Flap 1990; Zeggelink 1995). People interact with contacts at discrete times or maintain some relations while simultaneously cycling through others. The networks we inhabit, while perhaps maintaining islands of stability, are continuously changing. Even within relatively stable networks, our relational *activity* usually occurs at discrete points in time, as we visit relatives during the holidays or meet colleagues at lunch. At each interaction, we may (or may not) pass goods to our contacts, making it possible for them to later pass these goods to others.[2] How does this dynamic quality of social relations channel the flow of goods through a network? Can we identify systematic features based on relationship timing that limits diffusion potential within a particular network?

Past research on dynamic networks has focused either on the statistical problem of identifying change in the network or on substantive questions about what drives changes in network evolution. Statistically, if we assume that an observed network is a random realization of a true underlying network, we need to know whether the amount of change observed over time has been sufficient to conclude that the network has really changed. Under realistic data conditions, this work provides a metric that allows us to distinguish real change from sampling error and methods for modeling such change as a function of substantive features in the population (Leenders 1995; Snijders 1996, 1998, 2001; Snijders & Van Duijn 1997). Substantively, work often focuses on what creates change in a social network, ranging from relational stability at the dyad level to general tendencies in the global topology of the network (Doreian et al. 1996; Hallinan 1978; Hallinan & Tuma 1978; Hallinan & Williams 1987; Moody 1999; Suitor, Wellman & Morgan 1997).

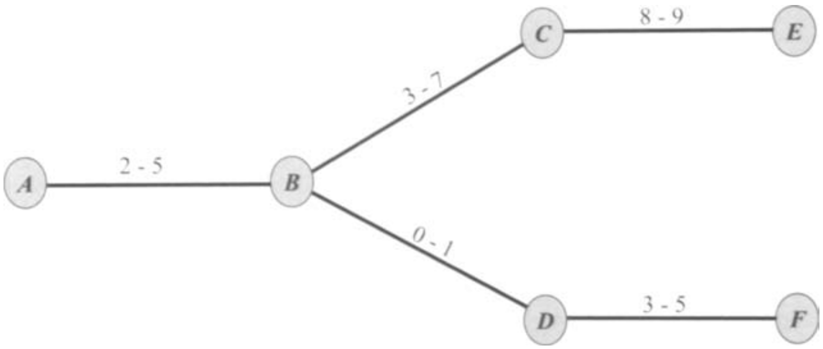
This work has shown repeatedly that social networks are fluid, changing in response to internal pressures (such as social balance) and external opportunity (such as classroom changes or tracking).

Though social diffusion is a decidedly dynamic process, most diffusion models rest on static networks, modeling the dynamic progression of a good through a fixed structure (Hedstrom, Sandell & Stern 2000; Kim & Bearman 1997; Myers 2000; Watts 1999b). That is, while there are elegent dynamic models of the flow of a good through a population that rest on an underlying network, the network *itself* is assumed to remain stable. Actors' states change over time, but the relations through which goods flow remain fixed.[3] In some cases, this simplification is inconsequential. If the network of interest is composed of durable, unchanging, and relatively constant relations (such as the geographic proximity of counties [Tolnay, Deane & Beck 1996]) or the opportunity to transmit a good through the network is continuous, then current approaches for identifying who can be reached through a network (and thus who is at risk from diffusion) are adequate. If, however, relations change quickly (relative to the transmission timing of the good) or network contact occurs at discrete moments, then any network measure based on the ability of one actor to reach another needs to account for this dynamism.

In this article, I explore diffusion limits based on time-ordered relations. I show that relationship timing acts like a railroad switch: Goods are channeled along different relational tracks depending on their sequential order. This article focuses on the logical limits imposed by timing, setting aside questions about network sampling and difficulties of collecting time-ordered network data. While these are very real concerns, since the constraints identified here must be active in the underlying network, I hope that this work will form a foundation for future statistical diffusion models. The insights presented below raise difficult theoretical problems for understanding reachability and thus many of the interesting properties of social networks. While the constraints that relational order place on diffusion are general, the examples I use are based on disease flow through intimate networks, such as needle exchanges or sexual activity. This is an ideal setting for identifying the implications of relational timing for diffusion, since sexually transmitted diseases (STDs) must travel through a connected network, providing a case of network diffusion that is not contaminated by nonnetwork sources.[4] Sexual networks are also dynamic relative to the life of pathogens traveling through them. Relationships start and end at different times, and one can be infected with a disease that was introduced to the network through a relation that has long since ended.

As with diffusion in general, most work on STD diffusion shows that network structure shapes the spread of disease. Centrality, for example, increases the likelihood of being infected (Bell, Atkinson & Carlson 1999; Morris 1993, 1997; Rothenberg et al. 1995). Disease epidemics tend to progress faster, but reach fewer people, when people have partners with similar numbers of partners (Anderson, Gupta & Ng 1990; Dietz & Hadeler 1988; Grady et al. 1997; Morris 1993, 1997

FIGURE 1: A Hypothetical Time-Ordered Network



but see Kault 1995 for a different view). Disease persistence in the population at large is credited to the existence of infected cores, which have been shown to be composed of a small number of highly connected actors who push infection into the wider population (Friedman et al. 1997). In much of this work, however, relationship timing is either ignored or implicit — for example, an accidental result of rules guiding relationship duration in a simulation.[5] The most influential explicit treatments of relational dynamics for diffusion rests on concurrency, which shows that when people have relationships that overlap in time (a wife and a mistress, for example), diffusion increases substantially (Morris & Kretzschmar 1995, 1997). Below, I show that concurrency marks one end of a diffusion-potential continuum defined by relationship order.

Building Intuition: Reachability in Dynamic Networks

Figure 1 presents a connected six-node, time-ordered network. Points represent actors and lines between points represent relationships. The numbers along the lines indicate the starting and ending dates of each relationship. For example, person *A* had a relationship with person *B* that started at time 2 and ended at time 5.[6] Person *B* is involved in three relationships, the first one with person *D* (started at $t = 0$, ended at $t = 1$) the second with *A* (started at $t = 2$, ended at $t = 5$), and the third with C, which overlaps the relation with *A* (started at $t = 3$, ended at $t = 7$). The graph in Figure 1 (ignoring time) is the kind of image one would get from standard network questionnaires asking respondents to list their interaction partners and accurately reflects the adjacency structure of the network. While we have a well-developed set of tools for analyzing this adjacency structure, ignoring time leads to substantive errors in our understanding of the diffusion potential represented by the network.

Most social network analysis packages would treat Figure 1 as a single connected network.[7] An analyst approaching this structure with standard tools would be interested in identifying central actors (node B), structurally equivalent positions (such as nodes D and C), or indirect path lengths (4 steps from actor *F* to actor *E*). From this perspective, the graph in Figure 1 is *connected* and all pairs are mutually reachable. Diffusion models usually assume that goods are free to flow throughout such a connected network. If there is an adjacency chain (a *path*) connecting two nodes, then a good should be able to flow between them.

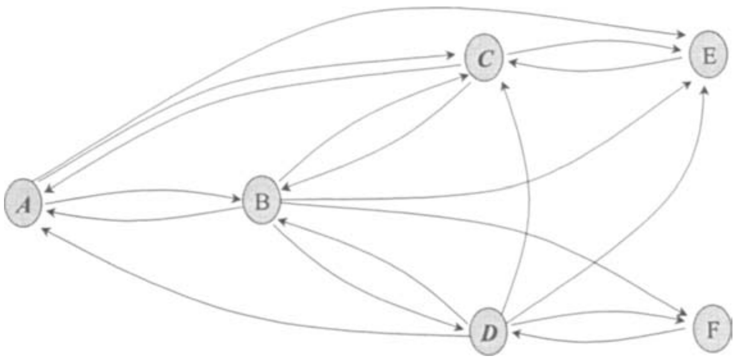
From a diffusion perspective, however, many of the apparent paths evident in this adjacency structure cannot transmit goods. Consider, for example, nodes *A*, *B*, and *D*. Because the *B-D* relation occurs prior to the *A-B* relation, *D* can send a good to *A*. *A*, however, cannot send a good to *D*, since the *A-B* link needed to connect them occurs after the *B-D* relation ended. Consider now the set of relations occurring among actors *A*, *B*, and *C*. Because *B* is concurrently related to both *A* and *C*, time-ordered paths from *A* to *C* and from *C* to *A* exist.

To capture the diffusion potential in a time-ordered network, we must identify all paths between pairs that could carry flow. Since a good cannot flow through a relation that has ended, we can identify potential diffusion paths as sequences of edges in which each new edge follows past edges in time. Walking through Figure 1, and tracing only those paths that follow each other in time, we get the *path graph* presented in Figure 2. Here, a curved arrow connects two nodes if there is a sequence of edges connecting the two nodes ordered such that a good could flow from the node at the base of the arrow to the node at the end.

In this path graph, all *direct* edges are transformed into two directed arcs (since any two actors who are directly connected could pass a good to each other) and every possible *indirect* relation is drawn as a directed arc if the good could flow through the path. For example, the indirect relation from *D* to *A* (through *B*) is shown by the curved directed arc from *D* to *A*. Because the *B-D* relation occurred before the *B-A* relation, *A* cannot infect *D*, and no arc is present. *B*'s concurrent relationships generate the two directed arcs between *A* and *C*. The arrows in Figure 2 exhaust all possible diffusion flows in this network. So, for example, the absence of an arc from *E* to *F* indicates that it is impossible for *E* to infect *F* (because *E*'s relation with *C* occurred after *C*'s relation with *B*).

All network models that pool relations over time implicitly assume a completely concurrent network. This assumption implies the greatest possible diffusion potential. Since actual diffusion paths are a subset of all possible paths in the collapsed network, any measure calculated on the collapsed graph that rests on indirect paths will be incorrect and likely misleading. For example, if time is ignored, each person in Figure 1 could reach every other person and reachability

FIGURE 2: Path Graph Derived from Figure 1



(the density of paths) would be 1.0. However, the actual reachability of Figure 1 is 0.6, or 40% lower than is implied by the collapsed graph.

The effect of timing on reachability is clearest when one compares graphs that differ only in the time order of relations. In such cases, reachability is affected only by the order of relations. Consider, for example, the five networks in Figure 3.

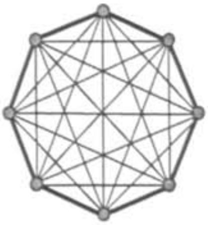
The networks in Figure 3 are all 8-node cycles, with direct relations indicated by heavy dark lines. The timing of each relation is labeled on the contact edges. Possible indirect paths implied by the contact network are presented with thin lines internal to the circle. Mutual paths have no arrows, asymmetric paths have arrows indicating the direction. While all five graphs in Figure 3 have the same adjacency structure, the relations occur at different times. In the top graph, all relations occur at the same time, and the path graph is the same as the collapsed graph. Since the underlying graph is a cycle, each person can reach every other person through 2 independent pathways (moving clockwise or counterclockwise through the graph). The reachability matrix in the second column shows this by indicating 2 paths between actors in every pair in the network, and the average number of paths between members of all pairs in the network is 2.

In the remainder of the graphs, there is some temporal variation on the edges, which leads to a significant change in the level of reachability. In the first case, relations are ordered in a long sequence. This results in an asymmetric contact structure, with one person able to reach (and be reached by) every other actor, while others can reach people who cannot reach them. It is instructive to compare network 2 with network 4, which differ only in the timing of a single relation (the time 5 relation changed to time 1). This one change lowers reachability by 36%,

FIGURE 3: Implications of Relation Order on Five Identical Contact Graphs

Network | Path Graph ($k[P]$) | Reach. $\Sigma(k[P])/n^2$

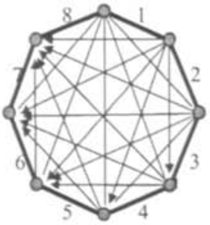
Concurrent



2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2

2.0

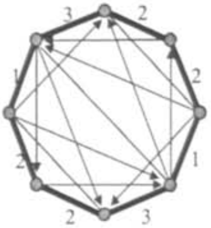
Serial Monogamy (a)



1	1	1	1	1	1	1	2
2		1	1	1	1	1	1
1	1		1	1	1	1	1
1		1		1	1	1	1
1			1		1	1	1
1				1		1	1
1					1		1
1						1	

0.67

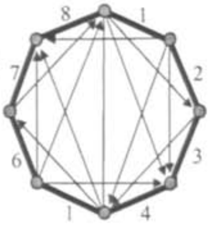
Mixed Concurrent



	1	1					1
1		1					1
1	1		1	1			1
1	1	1		1			1
			1		1	1	
			1	1		1	
1			1	1	1		1
1			1	1	1	1	

0.5

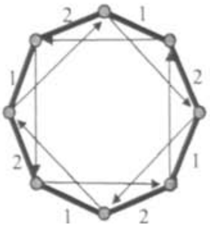
Serial Monogamy (b)



	1	1	1	1			1
1		1	1	1			1
	1		1	1			
		1		1			
1			1		1	1	1
1				1		1	1
1					1		1
1						1	

0.44

Minimum Reachability



	1	1					1
1		1					1
	1		1	1			
	1	1		1			
			1		1	1	
			1	1		1	
1					1		1
1					1	1	

0.38

showing that very minor changes in the network can have dramatic changes on reachability (a finding that mirrors Watts 1999a [p. 522] on the effect of minor topological changes). Graph 3 shows a network ordered somewhere between fully serial and completely concurrent, with four of the eight observed relations being concurrent (those happening at time 2). This results in two regions of the network, surrounding the concurrent relations, where flow can circulate through mutual infection paths. The final graph, 5, presents a network with no concurrent relations timed to weave between each other, which minimizes the diffusion potential. The set of five networks thus spans the range of maximum (1) to minimum (5) diffusion potential. Substantively, these small-network examples demonstrate how simply knowing the contact structure is insufficient to identify reachability, and thus potential diffusion, in networks composed of relations that occur at discrete times.

Formalizing Reachability in Time-Ordered Networks

The properties of time-ordered paths limit reachability in the network. Unlike untimed symmetric networks, indirect paths in a time-ordered symmetric network are neither necessarily mutual nor transitive. To develop tools for understanding diffusion potential in such networks, we need to formalize our representation of the network beyond simple images such as those presented in Figures 1 and 2. Following standard graph theory (Harary 1969), I represent the underlying adjacency structure as a graph, G(V,E), where actors are represented by V, the set of |v| vertices, and relations by E, the set of edges. I treat all edges in the graph as symmetric ($e_{vw} = e_{wv}$), and we say that an actor v is *adjacent* to actor w if $(v_v, v_w) \in E$.[8] A *path* in the network is defined as an alternating sequence of distinct nodes and edges, beginning and ending with nodes, in which each edge is incident with its preceding and following nodes. Actor v can *reach* actor w if there is a path in the graph starting with v and ending with w. Two paths from v to w are *node-independent* if they have only nodes v and w in common. If there is a path linking every pair of actors in the network, then the graph is *connected*. In general, a set is *maximal* with respect to a given property if it has that property but no proper superset does. A *component* of a graph G is a maximal connected subgraph of G. A clique is a maximal subgraph of G in which every pair of actors in the subgraph is adjacent.

We can build time into the network by defining starting and ending functions for the edges. Define $s(e_{vw})$ as the start time for the edge connecting v to w if $e_{vw} = 1$ and $f(e_{vw})$ as the finish time for the same edge.[9] These functions result in a valued graph, $G(V, E, s, f)$, where s and f are functions from E to the time scale. Two incident relations are *concurrent* if they exist at the same time, that is, if $s(e_{wm}) \leq s(e_{mv})$ and $f(e_{wm}) \geq s(e_{mv})$. A vertex, v, is *temporally reachable* from w if a properly time-ordered path starting with *w* and ending with *v* exists. Intuitively, a sequence of relations is properly time-ordered if each relation in the sequence precedes (or overlaps with) the other in time. Formally, one can define a one-step time-ordered path from *w* to *v* as p_{wv}^{1}; and use s(p_{wv}^{1}) and *f*(p_{wv}^{1}) to denote the corresponding start and finish times, which indicate the earliest and latest times for a possible event of transmission from *w* to *v*. Define longer paths recursively using

$$\begin{aligned} p_{wv}^{k+h} &= 1 \; if \;\; s\left(p_{wm}^{k}\right) \le f\left(p_{mv}^{h}\right), \text{ for some node } m \\ &= 0 \text{ otherwise} \end{aligned} \tag{1}$$

The length of the path is $k + h$, and we can define starting times based on our understanding of the goods flowing through the network. For goods that remain transmittable so long as an actor has them (such as HIV and many types of information), we would define the starting and ending dates of a particular path as

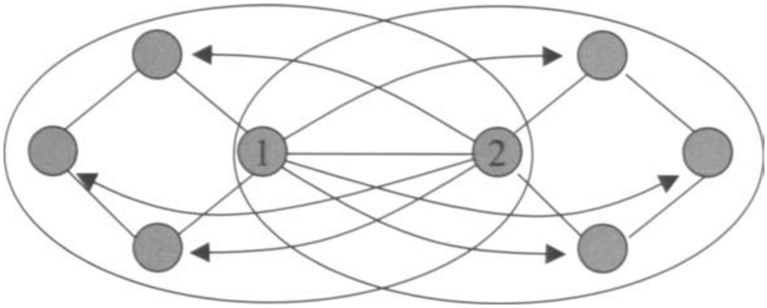
$$\begin{aligned} S_{m}\left(p_{wv}^{k+h}\right) &= \max\left(s(p_{wm}^{k}), s(p_{mv}^{h})\right) \\ f_{m}\left(p_{wv}^{k+h}\right) &= f\left(p_{mv}^{h}\right) \end{aligned} \tag{2}$$

That is, person *v* is at risk from a disease introduced by *w* along the path through *m* for a period starting at the latest value along the path and ending at the end of the *m*,*v* relation.[10] A *path graph*, P(*p*, V, *k*, *l*), can be defined as a directed graph derived from G(E, V, *s*, *f*) in which the new set, p, of edges consists of all (ordered) temporally reachable pairs in G, with P_{vw}=1 if there is a time-ordered path from *v* to *w* in G. Again, two functions further define P, with *k*(P_{vw}) indicating the number of paths from *v* to *w* and *l*(P_{vw}) indicating the length of a path from *v* to *w*. In substantive terms, P represents a *possible* path for diffusion between nodes, while *k* and *l*, in conjunction with the characteristics of the risk or resource flowing through the network, determine the *probability* of transmission.

Network Measurement Implications

Many structural network properties rest, explicitly or implicitly, on reachability in the network. To the extent that any of these properties is relevant for network diffusion, such measures will be inaccurate when used on a time-ordered network. This is clear with reachability, but is also true for any other network property that rest's on indirect connection. Many centrality measures that have been shown important for diffusion, such as betweenness centrality, information centrality, and closeness centrality, rest on the path structure of a network (Bell, Atkinson & Carlson 1999; Freeman 1978-79; Friedkin 1991; Rothenberg et al. 1995;

FIGURE 4: Reachability Component Structure Implied by Relationship Timing



Carlson 1999; Freeman 1978-79; Friedkin 1991; Rothenberg et al. 1995; Wasserman & Faust 1994). Betweenness and closeness centrality, for example, require identifying the shortest paths between pairs of actors in a network. When paths are limited by relationship timing, the set of shortest paths changes.[11] Information centrality, which has been shown to be correlated with disease diffusion in the STD context, rests on powers of the adjacency matrix and thus implicitly on the underlying path structure of the network (Rothenberg et al. 1995). In most cases, since actors in the collapsed graph bridge pairs that are, in reality, unreachable, any flow-based centrality measure calculated on the collapsed network will overestimate centrality. Given the nonlinear effects on path structure for small changes in relationship timing, this overestimation is likely highly sensitive to the particular timing in the network, and not simple random measurement error.

Topological measures that rest on connectivity in the network, such as the connectivity structure of the graph (White & Harary 2001), must trace the number of paths that link any two actors in a network. When time order makes traversing some paths impossible, the component structure changes. In an undirected graph, reachability is mutual: if *i* can reach *j*, then *j* can reach *i*. The asymmetry induced by relationship timing, however, means that undirected relations give rise to a type of weak component (*i* can reach *j*, but *j* cannot reach *i*), and the potential flow within these weak components overlap in complex ways. Consider, for example, Figure 4.

Figure 4 represents part of a path graph in which $e_{1,2}$ precedes all other relations in the network. Under this circumstance, person 2 can reach the people following person 1 indirectly, who may be able to reach each other, but they cannot reach node 2 (or any node following 2). Potential diffusion components will always contact network. Similarly, the flow of a good around an early relation has the separating structure (goods from node 2 flow to the left, while goods from node 1 flow right) evident in Figure 4. Empirically, the first step for identifying the importance of time-ordered paths in any given network is being able to identify all such paths, which requires tracing paths through a network. A new algorithm for doing so is presented in the Appendix.

Identifying Diffusion Limits

The basic requirement for diffusion is that nodes with a good be connected to those who do not have the good. Given that a connecting path exists, other properties, such as distance or location in a set of multiple paths may also be important, but the simple volume of ties in the network is a necessary minimum requirement for diffusion. For any such volume measure, it is useful to have an anchor that allows us to know whether the observed value is large or small. Below, I identify the minimum number of possible diffusion paths in a graph based on actor degree. This both provides a baseline for evaluating observed diffusion potential and demonstrates some of the basic reachability limits of time-ordered networks. Substantively, relative reachability, the ratio of observed paths to the minimum possible number of paths, ought to have a direct relation with diffusion level in the network. Networks with greater relative reachability ought to have higher levels of observed diffusion, since there are a greater number of potential pathways through the network.

Define relative reachability, R, as

$$R = \frac{P_{++}}{\min(P_{++})} \tag{3}$$

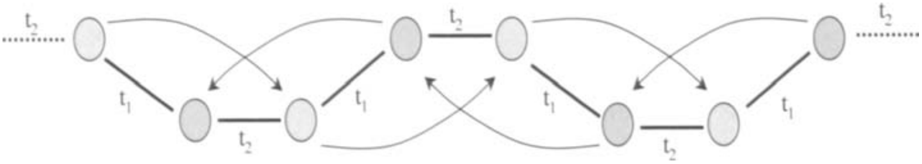
where P_{++} is the number of indirect paths observed in the network and $\min(P_{++})$ is the minimum possible number of paths in a graph of the same contact structure. An analogous measure for individuals is simply the number of paths to (from) the node divided by the minimum possible number of ties to (from) ego. That is,

$$R_{i\bullet} = \frac{P_{i+}}{\min(P_{i+})} \tag{4}$$

$$R_{\bullet i} = \frac{P_{+i}}{\min(P_{+i})} \tag{5}$$

where $\mathrm{R}_{i.}$ is the number of people whom ego can reach and $\mathrm{R}_{.i}$ is the number of people who can reach ego. The key to any such measure, however, is to identify min(P), the minimum possible reachability in a network.

FIGURE 5: Minimum Path Graph for a 2-Regular Contact Graph



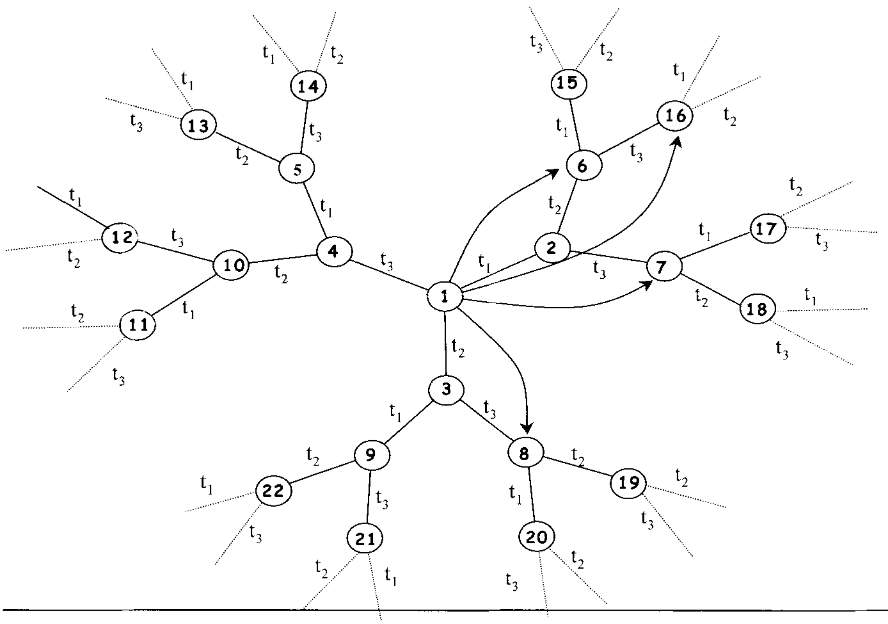
Minimum Reachability Graphs

Given a connected component, how might relations be ordered in time to minimize the number of indirect paths, and thus diffusion, in the network? Concurrent partnerships create implied relations that are symmetric among members of the concurrent set, adding two arcs to the path graph for every pair of actors concurrently linked. Thus, any *minimum* path graph must contain no concurrent relations. This simplifies our presentation, as strictly nonoverlapping relations can be thought of as existing in discrete time periods.[12] I start by focusing on graphs that have no cliques.[13] I identify minimum reachability by identifying the reach implications for specific substructures and then expand these substructures to an arbitrarily large graph. While exact formulas can be identified for the regular graphs presented below, the graphs themselves are unrealistic. As will become clear, however, the minimum number of paths in a given network depends more on degree than on the global topology. Contact structure limits *who* is reached through a path, while degree determines the *number* of possible paths, and thus the unrealistic structures presented below are not as limiting as they might seem.[14]

The number of contacts any given node eventually reaches is equal to the number of future partners his or her partners have. If ego's partner has only one other (future) partner, then ego's maximum indirect reach is one additional person. Consider the graph presented in Figure 5. This graph is 2-regular — each person has two (and only two) contacts.[15] If we weave time periods between pairs of actors, such that every path of length three has the pattern: "early, later, earlier" then each person will be able to reach only one other person indirectly.[16] This graph has the lowest possible reachability for a connected 2-regular graph of *g* nodes.

The minimum number of lines of this graph can be calculated as follows: for the single line graph (2-regular except for the ends), each nonending person has

FIGURE 6: Minimum Path Graph for a 3-Regular Tree



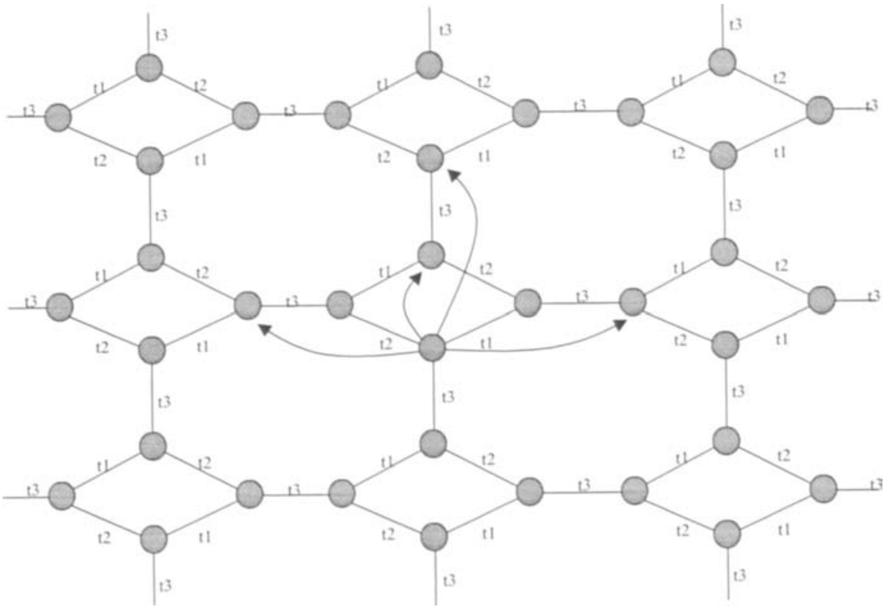
two direct contacts, generating $2(g-2)$ arcs. Each nonending person contributes 1 additional indirect arc, creating another $g-2$ arcs to the total. The two people at the ends of the chain (assuming a t_2 end) will have one real relation, and no implied relations, generating two arcs total. Thus, the total number of lines in the graph will be

$$\begin{aligned} l &= 2(g-2) + g - 2 + 2 \\ &= 3g - 4 \end{aligned}$$

The problem is simplified if instead of a single line, we had one large cycle (by, for example, putting a t_1 relation between the ends of the graph in Figure 5). In the cycle case, each person would have two direct ties and one indirect tie, resulting in a minimum of 3g lines.

We can extend this idea to graphs with higher degree by extending the notion of an interwoven sequence of relations across all relations in the graph. For a 3-regular graph, the number of possible adjacency structures is large, and I focus on two classes of 3-regular graphs below: those that allow complete triads and those that do not.[17] In both, I require that every person have three relations at three interwoven, nonoverlapping times. To weave time periods in a 3-regular graph, every node must be involved in a relation at every time, and only one relation in

FIGURE 7: 3-Regular Graph with Four and Eight Step Cycles



have a relation in each of the three time periods. Such a graph is presented in Figure 6.

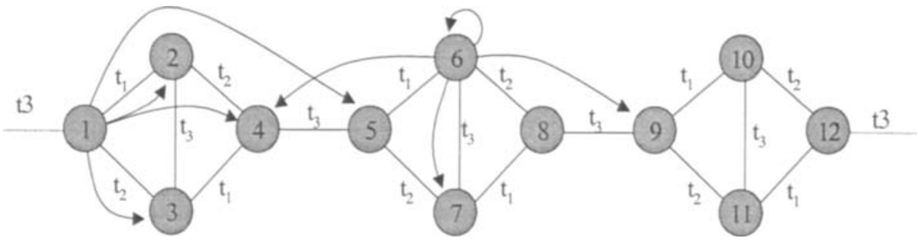
This graph is an ever-expanding spanning tree. The person at the center has three partners, each of whom has three partners.[18] If we enumerate the set of possible time-ordered paths in this network, we find four possible path types (labeled by the time periods of the relations):

(t_1: t_2) (e.g.: 1 → 2 → 6)
(t_1: t_3) (e.g.: 1 → 2 → 7)
(t_2: t_3) (e.g.: 1 → 3 → 8)
(t_1: t_2: t_3) (e.g.: 1 → 2 → 6 → 16)

As it turns out, there is only one position in this network, as the graph looks the same from every point in the graph (assuming the ends continued indefinitely). Thus, every ego will start each of these four paths once, meaning that each person has 4 indirect paths in addition to their 3 direct arcs, for a total of 7*g* arcs.

Interestingly, the weaving of three time periods results in four indirect arcs, *regardless of the contact pattern.* The pattern of ties will affect *whom* the indirect arcs reach, but not the number of paths. In some cases, indirect paths will reach

FIGURE 8: 3-Regular Path Graph with Small Cycles



new nodes, representing a case where infection could move through a wider population. In other cases, indirect arcs will provide alternative pathways to the same nodes. To see this clearly, compare the graph presented in Figure 7 to that in Figure 8. In Figure 7, we have a 3-regular graph arranged in a grid with many small cycles. Small groups of four are linked through t_1 and t_2 relations, and each of these four-person groups is linked to others through t_3 relations. All actors are equivalent with respect to their potential to infect others: each person can reach four other people. These indirect relations are indicated with curved arrows for one node in the graph.

In contrast, Figure 8 allows complete triads. Like the graph in Figure 7, the basic structure in Figure 8 revolves around small groups of four nodes. In Figure 7, all t_3 ties link new clusters. In Figure 8, one t_3 relation links a pair within the foursome (e.g., the link between actors 2 and 3). Because of the interwoven time periods, each actor still sends four indirect arcs. However, because of the small cycles, many of these indirect links relink already connected actors. Two types of positions emerge from this graph. Those nodes that bridge the four-member clusters (nodes 1 and 4, for example) are connected with two new people through indirect paths (e.g., 1→2→4 and 1→2→4→5) and reconnected to two others (e.g., 1→3→2 and 1→2→3). Nodes with all direct ties internal to the group also reach two new people (e.g., 6→8→9 and 6→5→4), and relink to one other node and themselves (6→5→7 and 6→5→7→6 respectively). These two positions are drawn in Figure 8 for nodes 1 and 6 respectively.

The total number of indirect relations is still 4*g* for a graph of the type in Figure 8, but the number of *new* people contacted would only be 2*g*. On the one hand, we would expect diffusion to progress slowly from such clustered settings, as there are few new people contacted. On the other hand, once introduced to one of these clusters, there are many opportunities for circulation.

Other 3-regular patterns that mix closed triads and trees could be proposed, and the method for identifying a minimum path density would still be the same. The previous examples show, however, that there is a complex interaction between relationship timing and relational pattern. In all these examples, the networks are identical from ego's standpoint: Every person is involved in three nonconcurrent relations with someone else who was also involved in three relations. In each case, the number of indirect relations was constant and can be identified by exhausting the number of time sequences in the network ($t_1 \rightarrow t_2 \rightarrow t_3$ etc). The relationship *order* governs the number of indirect ties, while the contact *structure* governs whether paths are new links or reconnections. This finding implies that reachability rests on the adjacency structure of the edges, a finding foreshadowed by Kretzschmar and Morris's (1996) finding that concurrency can be measured through the connectivity structure of edges.

In general, the minimum number of time-ordered paths through a network with regular degree will be achieved by ensuring only asymmetric paths and minimizing the length of all such paths. We can prevent symmetric paths by forbidding concurrent relations, which in turn requires that we set the number of time periods equal to the degree and then arrange them such that every person is involved in one relation at each time. We minimize the length of all such ties by arranging them such that sequences of relations are interwoven. Assuming such an ordering exists for the structure in question, then the number of paths each person is involved in could be expressed as follows:

$$P_i = \sum_{l=2}^{T} \frac{T(T-1)(T-2)\ldots(T-l-1)}{l!} \qquad (6)$$

where T is the number of time periods and equals the degree of the graph, and l is the length of the given path, which can range from 2 to $\underline{T}$. For $T = 3$, there will be $3(2)/2 = 3$ indirect paths of length 2, and $3(2)(1)/(3)(2)(1) = 1$ path of length 3 for 4 total per person, and thus $4g$ indirect paths per graph.

These minimum values provide a guidepost for observed graphs, in that we can compare the observed number of indirect relations to the minimum graph with similar direct degree. Of course, the usefulness of this minimum value in an index of indirect relational volume rests on how well the assumptions of a regular graph fit with empirical settings. More work is needed to determine if the minimum indirect graph for a network in which the average degree is close to the minimum distance for a k-regular graph, or how various patterns of mixed degree nodes affect the number of indirect links.

If one is uncomfortable with the regularity assumption, an algorithmic approach can be used to simulate the minimum reach for any network structure. One would simply use TIME_BFS, as presented in the Appendix, to identify long indirect paths and change the timing of a relation in the middle of that path to an early date (say $t_{min} - 1$) to break the indirect path. Repeat this process for all paths longer than some specified length (say 3), and one would achieve a minimum ordering of relations for the graph. Some caution is needed, however, because the same edge functions differently in different paths. Changing the time of one relation to shorten a given path may inadvertently lengthen another, and thus one cannot prove that such an algorithm would converge on a true minimum for the given structure.

An Empirical Example

Though it is tautological that temporal ordering constrains reachability within networks, to what degree are real networks temporally bounded? Below I present data from the National Longitudinal Survey of Adolescent Health (Add Health) that demonstrates how timing partitions an observed network component into smaller risk pools.

THE ADD HEALTH ROMANTIC NETWORK

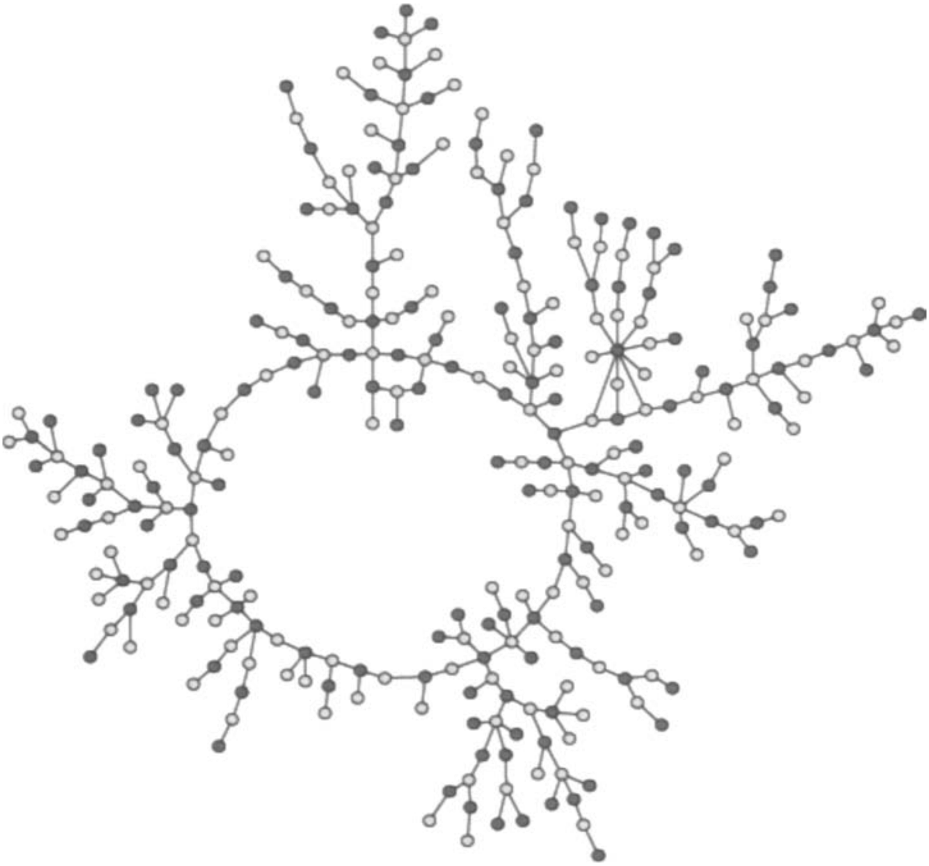
Add Health collected romantic relationship data from more than 20,000 American adolescents. In two large settings, Add Health collected romantic relationship data from all students in the school. The example I use below comes from "Jefferson" high school. Each student surveyed was asked to report on romantic and nonromantic sexual relations that were ongoing within the last eighteen months, which allowed us to construct an image of the school's romantic network (see Bearman, Moody & Stovel 2002 for details about the romantic network structure in Jefferson).

Figure 9 presents the largest component of the direct contact network, with males plotted in dark gray and females in light gray. Of the 832 students in Jefferson who completed the survey, 535 were involved in a romantic relation with at least one other person from the school. Of these 535 students, 288 were involved in the largest connected component. The longest path in this component (ignoring time) is 37 steps and 40 students are members of the largest bicomponent (the large cycle evident in the graph plus the 3 nodes in a small cycle at the top of the figure).[19]

Our vision of this network changes dramatically if we identify where disease could move based on the timing of relations. Figure 10 plots the same network with all direct relations removed and all indirect relations included. In this plot, each line marks a potential infection path. Grey arrows represent asymmetric paths, while dark lines represent symmetric paths. Pockets of mutually reachable students are evident as regions of high density in the path graph.

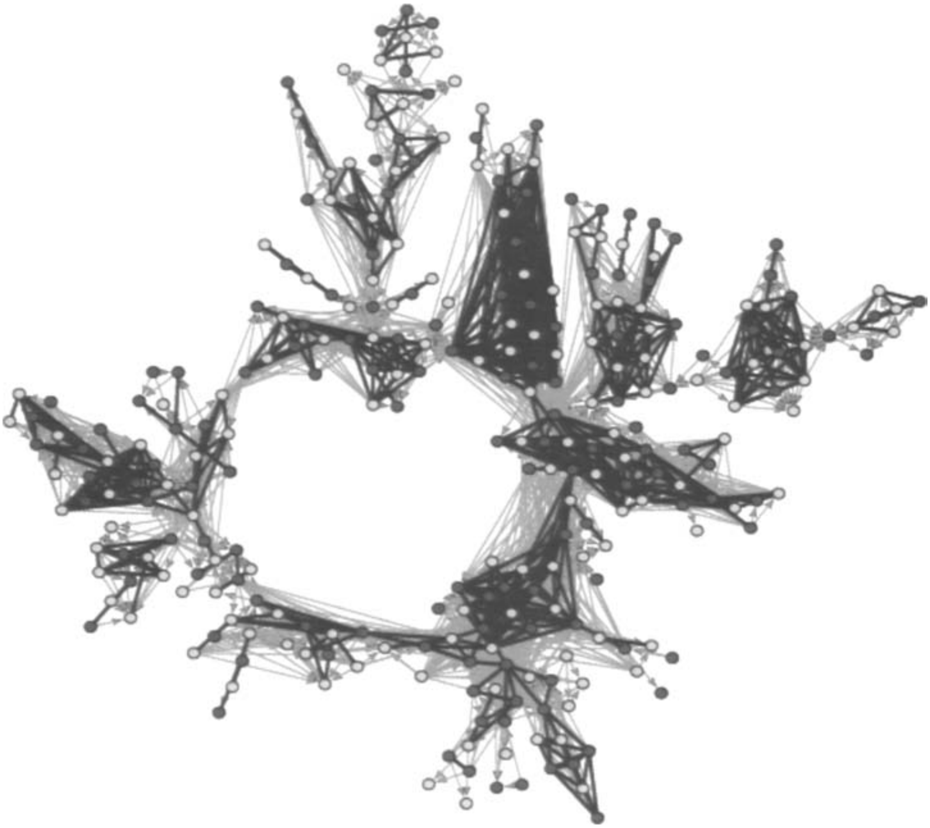
The graph in Figure 10 contains clear subsections of mutually reachable pairs, and the longest time-ordered path is 12 steps long. Thus, while the collapsed graph

FIGURE 9: Direct Romantic Relations in Jefferson High School



component has a diameter of 37 steps, the longest possible diffusion path is only 12 steps. The observed graph has 3,221 indirect relations, while the minimum possible graph[20] has 1,035 relations, for a reachability ratio (*R*) of 2.82, indicating that the number of potential infection paths in Jefferson school is nearly 2.8 times that of the minimum possible for a graph of the exact same contact structure.[21] In the observed graph, the person who could infect the greatest number of people could infect up to 40 others, while in the minimum graph the same person could infect only 14, indicating that his potential to infect in the observed network is just over three times the minimum possible.

FIGURE 10: Indirect Relations in Jefferson High School



Conclusion

Whether we are speaking of disease diffusion through sexual contact or information diffusion through conversation, transmission occurs at specific points in time and thus limits diffusion within a network. Timing bounds diffusion by limiting the paths through which risks and resources flow. Like switches on a railroad track, relationship timing routes network flow to particular subsets of the network. This occurs because nonconcurrent relations give rise to asymmetric diffusion paths, sending flow down the network equivalent of a one-way street. This time-ordered view of networks possess new measurement and theoretical challenges for diffusion research and opens a set of new analytic questions linking graph topography and timing to reachability in a social network.

First, network measures of centrality and subgroup structure that explicitly account for relationship timing are needed. Substantively, centrality within the diffusion context should measure the extent to which a particular actor is important for potential diffusion. The substantive meaning of importance will depend on the context but will likely relate to how many paths pass through a given actor, how many paths a given actor starts, or how close an actor is to any other actor in the network. Therefore, time-dependent centrality measures that make use of the number and length of time-ordered paths are obvious choices. The number of paths in the network, $k(\mathbf{P}_{wv})$, and the length of each path, $l(\mathbf{P}_{wv})$, provide the basic information needed to construct time-ordered measures analogous to betweenness, flow, or closeness centrality.

Identifying meaningful subgroups within time-ordered networks will require an explicit understanding of the content of network flow, but the distinction between strong and weak components will likely be of primary interest. A strongly connected component implies a maximal subgraph where every member of the subgraph can reach every other member. This is equivalent to a clique in **P**, and applying standard clique detection algorithms directly to **P** will induce strong components of the time-ordered graphs. These strong components are the places in the network where network flow can recirculate, creating pools of risks or resources. As with networks more generally, we can extend strong connectivity by identifying the number and length of paths through which pairs of actors are reachable (Moody & White 2001; White & Harary 2001). A strong component need only imply a single path between actors of every pair, while bicomponents (k-components) imply $2(k)$ paths between actors of every pair. Increased connectivity is an important factor in diffusion, as multiple paths increase the probability that diffusion occurs whenever dyadic transmission is uncertain.

A more difficult, though potentially theoretically interesting, case can be made for substantively interpreting and organizing weak connectivity within time-ordered networks. We can define each node in a network as a potential diffusion source with different potential sinks. That is, the set of all people downstream from actor w defines a region of the graph similar to a component, which are at risk from w. An important question becomes how these risk components connect as a function of relationship structure and timing. For example, a classic disease core suggests that a strongly connected component radiates disease to many potential sinks. But one can imagine scenarios where a diffuse set of sources converge on a single sink that has the result of exposing individuals to a broad selection of past network flow. Closely related, understanding how different mixing patterns generate risk components is a key question and will involve extending work on assortative mixing to work on relationship timing (see Morris & Kretzschmar 1995 for related work on concurrency and mixing pattern).

Both connectivity and centrality will be important for understanding how the *possibility* for diffusion defined by P relates to the *probability* of diffusion. The diffusion probability is bounded by the possible paths (i.e., = 0 if there is no path) and is a function of the pairwise transmission probability at each dyad ($p[t_{wv}]$). As long as the transmission probability is less than 1, the probability of diffusion between two nonadjacent nodes depends on the number and length of paths connecting the pair. In general, longer paths lower transmission probability, since the probability of traveling through any given path of length l is equal to t_{wv}^{l} . Conversely, the number of independent paths provides alternative diffusion routes and thus increases the diffusion probability.

Because diffusion depends as much on one's partner's future partner's behavior as one's own, this article shows clearly that diffusion is an emergent property of the network. The distribution of paths reflect global connectivity properties, insofar as there is no a priori way of computing path distributions from lower-order summaries of the networks. The challenge of explaining the emergent property reduces to understanding how local interaction combines to produce the global pattern, which, in this context, implies identifying how different temporal orders emerge in a network. What behavioral factors, for example, lead to long relationship chains rather than short relationship chains or multiple sinks versus diffusion cores? Given the known importance of assortative mixing patterns for diffusion, it is likely that explorations into the path structure under different mixing regimes has implications for reachability in closed populations.

This article also shows that there is a range of risk associated with relationship timing that has a minimum with interwoven, discrete relationships and a maximum with completely concurrent networks. From a practical standpoint, most analysts use a collapsed network that assumes a completely concurrent network and thus dramatically overestimate reachability in the network. Concurrency matters because indirect relations flow symmetrically through concurrent relations, allowing the maximum possible reach for any given actor. Neither of these poles is likely to be observed in large networks. While we currently have measures for concurrency at any point in time (Kretzschmar & Morris 1996), further work on the developmental history of networks ought to quantify risk below this maximum threshold.

Finally, though this article has used romantic networks to identify the effect of timing on diffusion, the implications hold for any network where resources or risks travel through the network at discrete times. A scientific coauthorship network, for example, has implicit timing on the edges defined by publication dates. Our ability to predict the rise of scientific innovation might be enhanced if we were to identify where paths within this network lead to information sinks. Similarly, networks of business elites constructed through overlapping directorates can be time-ordered by membership dates that might be useful in predicting company mergers or splits. Other applications abound, reaching across political, social, and disease networks. While social network researchers have started to focus increasing attention on the temporal aspects of networks (Doreian & Stockman 1996; Snijders 2001; Snijders & Van Duijn 1997; Suitor, Wellman & Morgan 1997; Weesie & Flap 1990), little work has looked at the implications of network change for the things that travel through networks. Practically, this work suggests that by collecting data on interaction timing, we can sharpen our understanding of diffusion in multiple contexts and increase the predictive power of our models.

Notes

1. I use *goods* here generically, as many of the items most often of interest to diffusion researchers, such as sexually transmitted diseases, are decidedly not good. The generic term saves us the awkwardness of *risks and resources* which is otherwise more accurate.

2. This fact distinguishes social networks from other networks, such as electric power grids, highway systems, or computer networks, where goods may flow through the network continuously. It is reasonable to assume that electricity or traffic can flow through their respective networks in conformity with only the adjacency structure. Social networks, in contrast, are not continuously connected. Instead, actors use relations to carry goods through the network at particular points in time. For these reasons, the observed pattern of durable relations represent all possible interactions, many of which are never activated to transmit a particular good through the network.

3. When simulation models include "removed" in the states of actors in the network, the underlying graph can change through the removal of arcs (Watts 1999). Note that this occurs as a *consequence* of diffusion, however, and is not analyzed explicitly as an element that *shapes* diffusion.

4. The potential for disease flow is shaped jointly by the structure of the contact network and the probability of transmission along any observed edge in the network. This probability is highest for unprotected intercourse, dropping significantly for other types of fluid exchange. As such, the observed contact structure can be thought of as a worst-case field for disease transmission.

5. One of the troubling aspects of many such simulations is the large within-simulation diffusion variation (Morris & Kretzschmar 1995), which often dwarfs mean differences between simulation settings. At least part of this within-simulation variance is likely due to factors associated with relationship timing.

6. I treat time periods as discrete. If v has a relation with w that starts at time 1 and ends at time 2, and v has a relationship with q that starts at time 2 and ends at time 3, I consider the two time-2 relations as overlapping. I also assume the relationship is ongoing between the start and end date. For a model with different discrete time assumptions, see Helander and Batta 1994.

7. PAJEK (Batagelj & Mrvar 2001) is the only program I am aware of that allows one to work with networks having timed edges. All procedures described in this article are available from the author as SAS programs. Of course, one would not need a package

to "analyze" a network as small as that presented in Figure 1; it is simply an example of the types of networks people collect.

8. I assume that actors do not relate to themselves and thus $(v_i, v_i) \notin E$.

9. If $e_{v,w} = 0$, then $s(e_{v,w})$ and $f(e_{v,w})$ are undefined. In the rare case that v and w have multiple start times, under most circumstances it would likely be acceptable to collapse $s(e_{v,w})$ to the earliest start time and $f(e_{v,w})$ to the latest time. The notation presented above differs from that used in statistical models for dynamic networks. Such models rest on Markov graph models and treat the network as an adjacency matrix, **X**, that is subscripted over time ($\mathbf{X}_t$). While this is convenient for linking changes in the graph to Markov chain models, the notation is less than ideal when dealing with goods flowing through the network. Because paths in time-ordered graphs are neither necessarily transitive nor symmetric, standard matrix multiplication methods for identifying paths (Wasserman & Faust 1994) are not possible. Instead, one must search all paths recursively. The notation I use here facilitates developing algorithms that are based on such searches through the network. Moreover, by treating time as a function on the edges, we can easily build in good-specific properties, such as windows of infectability for particular diseases, which would be particularly cumbersome with some other notation.

10. While I do not pursue it in this article, it is obvious that we can sharpen the path definition in equation 2 with respect to various flow properties. For example, if a disease is infectious for a known window of time, then equation 1 can be defined such that only new links that fall within the window lead to paths, and equation 2 is correspondingly bounded by the infectious window of the particular disease.

11. Note that the substantive importance of shortest path in a time-ordered network is somewhat problematic. Do we focus on the number of steps or the difference in time between the start of the path and the end of the path? The choice likely depends on the specific good flowing through the network. Information, for example, may be more valuable if received quickly, while diseases with a transmission probability less than 1 are more likely to travel through paths of few intermediaries.

12. If every person has only one partner, then indirect connections are impossible and the problem of identifying minimum diffusion potential is trivial. Substantively, we are interested in identifying the structures that *potentially* link an unlimited number of people, and I thus require every actor to have at least two relations. The indirect connection problem is also trivial if every node is adjacent to every other node (a complete clique), and thus I focus on graphs that are neither trivial nor complete.

13. The prohibition of connected cliques is an important distinction in the sexual network context, differentiating heterosexual from homosexual networks. Because heterosexual networks are bipartite graphs, it is impossible to create complete triads, which are the foundation of connected cliques.

14. Additionally, if one can identify a set of equivalence positions on an observed network, this expansion procedure could fit bounds for more complex graphs by fitting bounds to each part of the network. Just as one can calculate the area of complex polygons by identifying simple subparts, one can identify the minimum reachability in an observed graph by identifying simple subsections of the network.

15. The graph is strictly 2-regular if it eventually curves in on itself and forms a loop. Else, the graph forms a single line, which is 2-regular for every node except the end nodes of the line, which have one partner each.

16. It is important to note that times such as "early" and "late" are not assumed constant across paths in the network. All that matters is the *order* of the relations in that particular sequence.

17. The maximum number of such patterns would be difficult to calculate, as one could mix elements of each of the substructures described below in potentially endless ways. The topographical variety of structure depends, however, on the number of substructures allowed in the network. By forbidding complete triads, we constrain the set of possible graph patterns considerably. See Johnsen (1985, 1986) for a discussion of similar constraints in friendship networks.

18. Again, the assumption is that this pattern continues indefinitely, or eventually loops in on itself — as would be the case if nodes 11 and 22 shared a t_3 relation.

19. It is important to note that there may be connections among students through people outside the school that these data do not capture. The analysis presented here identifies the extent of diffusion *within* this set of romantic relations. This kind of boundary problem is not unique to the processes identified here, however, and is evident in all network studies that must, for practical reasons, bound the network.

20. Based on the algorithmic procedure outlined above.

21. Students in the largest component have about two partners each. If this were a regular graph with degree 2, the implied number of arcs would be 3g, which would yield an R of (3,221/[3*288]) = 3.73. A regular graph of degree 3 would result in an *R* of 1.11. The regular graph count thus bounds the iterated in this case.

References

Anderson, R.M., S. Gupta, and W. Ng. 1990. "The Significance of Sexual Partner Contact Networks for the Transmission Dynamics of HIV." *Journal of Acquired Immune Deficiency Syndrome* 3:417-29.

Batagelj, Vladimir, and Andrej Mrvar. 2001. *PAJEK*. Version 0.71.

Bearman, Peter S., James Moody, and Katherine Stovel. 2002. "Chains of Affections." Manuscript, Columbia University.

Bell, David C., John S. Atkinson, and Jerry W. Carlson. 1999. "Centrality Measures for Disease Transmission Networks." *Social Networks* 21:1-21.

Burt, Ronald S. 1987. "Social Contagion and Innovation: Cohesion versus Structural Equivalence." *American Journal of Sociology* 92:1287-335.

Carley, Kathleen. M. 1991. "A Theory of Group Stability." *American Sociological Review* 56:331-54.

Coleman, James S., E. Katz, and H. Menzel. 1966. "The Diffusion of an Innovation among Physicians." *Sociometry* 20:253-70.

Dietz, Klaus, and K.P. Hadeler. 1988. "Epidemiological Models for Sexually Transmitted Diseases." *Journal of Mathematical Biology* 26:1-25.

Doreian, Patric, Roman Kapuscinski, David Krackhardt, and Janusz Szczypula. 1996. "A Brief History of Balance through Time." *Journal of Mathematical Sociology* 21:113-31.

Doreian, Patrick, and F. Stockman. 1996. *Evolution of Social Networks*. Gordon & Breach.

Freeman, Linton C. 1978-1979. "Centrality in Social Networks." *Social Networks* 1:215-39.

Friedkin, Noah E. 1991. "Theoretical Foundations for Centrality Measures." *American Journal of Sociology* 96:1478-504.

Friedman, Samuel R., Alan Neaigus, Benny Jose, Richard Curtis, Marjorie Goldstein, Gilbert Ildefonso, Richard B. Rothenberg, and Don C. Des Jarlais. 1997. "Sociometric Risk Networks and HIV Risk." *American Journal of Public Health* 87:1289-96.

Grady, William R., John O.G. Billy, Daniel H. Klepinger, and Anjanette Nelson-Wally. 1997. "The Sexual Relationship Behavior of Unmarried Men." Unpublished manuscript, Batelle Memorial Institute.

Hallinan, Maureen T. 1978. "The Process of Friendship Formation." *Social Networks* 1:193-210.

Hallinan, Maureen T., and Nancy B. Tuma. 1978. "Classroom Effects on Change in Children's Friendship." *Sociology of Education* 51:270-82.

Hallinan, Maureen T., and Richard A. Williams. 1987. "The Stability of Students' Interracial Friendships." *American Sociological Review* 52:653-64.

Harary, Frank. 1969. *Graph Theory*. Addison-Wesley.

Hedstrom, Peter, Rickard Sandell, and Charlotta Stern. 2000. "Mesolevel Networks and the Diffusion of Social Movements: The Case of the Swedish Social Democratic Party." *American Journal of Sociology* 106:145-72.

Helander, Mary E., and Rajan Batta. 1994. "A Discrete Transmission Model for HIV." Pp. 585-611 in *Modeling the AIDS Epidemic: Planning, Policy and Prediction*, edited by E.H. Kaplan and M.L. Brandeau. Raven Press.

Johnsen, Eugene C. 1985. "Network Macrostructure Models for the Davis-Leinhardt Set of Empirical Sociomatrices." *Social Networks* 7:203-24.

———. 1986. "Structure and Process: Agreement Models for Friendship Formation." *Social Networks* 8:257-306.

Kault, David A. 1995. "The Impact of Sexual Mixing Patterns on the Spread of AIDS." *Mathematical Biosciences* 128:211-41.

Kerckhoff, Alan C., and Kurt W. Back. 1968. *The June Bug: A Study of Hysterical Contagion*. Appleton-Century-Crofts.

Kim, Hyojoung, and Peter Bearman. 1997. "Who Counts in Collective Action? The Structure and Dynamics of Movement Participation." *American Sociological Review* 62:70-93.

Kretzschmar, Mirjam, and Martina Morris. 1996. "Measures of Concurrency in Networks and the Spread of Infectious Disease." *Mathematical Biosciences* 133:165-95.

Leenders, Roger Th. A.J. 1995. "Models for Network Dynamics: A Markovian Framework." *Journal of Mathematical Sociology* 20:1-21.

———. 1996. "Longitudinal Behavior of Network Structure and Actor Attributes: Modeling Interdependence of Contagion and Selection." Pp. 165-84 in *Evolution of Social Networks*, edited by P. Doreian and Frans N. Stokman. Gordon & Breach.

Moody, James. 1999. "The Structure of Adolescent Social Relations: Modeling Friendship in Dynamic Social Settings." Ph.D. diss., Dept. of Sociology, University of North Carolina, Chapel Hill.

Moody, James, and Douglas R. White. 2001. "Structural Cohesion and Embeddedness: A Hierarchical Conception of Social Groups." Manuscript, Ohio State University.

Moore, E.F. 1957. "The Shortest Path through a Maze." Pp. 285-292 in *Proceedings of the International Symposium on Switching Theory,* part 2. Harvard University Press.

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):S209-16.

Morris, Martina, and Mirjam Kretzschmar. 1995. "Concurrent Partnerships and Transmission Dynamics in Networks." *Social Networks* 17(Special Issue: *Social Networks and Infectious Disease: HIV/AIDS*):299-318.

———. 1997. "Concurrent Partnerships and the Spread of HIV." *AIDS* 11:641-48.

Myers, Daniel J. 2000. "The Diffusion of Collective Violence: Infectiousness, Susceptibility, and Mass Media Networks." *American Journal of Sociology* 106:173-208.

Rothenberg, R.B., John J. Potterat, Donald E. Woodhouse, William W. Darrow, Stephen Q. Muth, and Alden S. Klovdahl. 1995. "Choosing a Centrality Measure: Epidemiologic Correlates in the Colorado Springs Study of Social Networks." *Social Networks 17* Special Issue: *Social Networks and Infectious Disease: HIV/AIDS:* 273-97.

Sanil, Ashish, David Banks, and Kathleen Carley. 1995. "Models for Evolving Fixed Node Networks: Model Fitting and Model Testing." *Social Networks* 17:65-81.

Smith, David A., and Douglas R. White. 1992. "Structure and Dynamics of the Global Economy: Network Analysis of International Trade 1965-1980." *Social Forces* 70:857-93.

Snijders, Tom A.B. 1996. "Stochastic Actor-Oriented Models for Network Change." *Journal of Mathematical Sociology* 21:149-72.

———. 1998. "Models for Longitudinal Social Network Data." Unpublished manuscript, University of Groningen.

———. 2001. "The Statistical Evaluation of Social Network Dynamics." *Sociological Methodology.* 31:361-395

Snijders, Tom. A.B., and Marijtje A.J. Van Duijn. 1997. "Simulation for Statistical Inference in Dynamic Network Models." Pp. 493-512 in *Simulating Social Phenomena*, edited by R. Conte, R. Hegselmann, and P. Terna. Springer.

Suitor, Jill. J., Barry Wellman, and David L. Morgan. 1997. "It's about Time: How, Why and When Networks Change." *Social Networks* 19:1-7.

Tolnay, Stewart E., Glenn Deane, and E. M. Beck. 1996. "Vicarious Violence: Spatial Effects on Southern Lynchings, 1890-1991." *American Journal of Sociology* 102:788-815.

Valente, Thomas W. 1995. *Network Models of the Diffusion of Innovations.* Hampton Press.

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

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

———. 1999b. *Small Worlds: The Dynamics of Networks between Order and Randomness.* Princeton University Press.

Weesie, Jeroen, and Henk Flap. 1990. *Social Networks through Time.* Utrecht, Netherlands: ISOR.

White, Douglas R., and Frank Harary. 2001. "The Cohesiveness of Blocks in Social Networks: Node Connectivity and Conditional Density." *Sociological Methodology* 31:305-59.

Zeggelink, Evelien P.H. 1995. "Evolving Friendship Networks: An Individual-Oriented Approach Implementing Similarity." *Social Networks* 17:83-110.

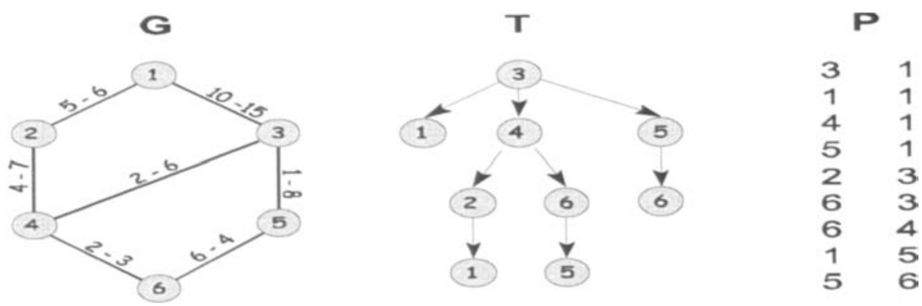
APPENDIX A: Identifying Indirect Paths in Time-Ordered Graphs

Many features complicate identifying reachable pairs in a time-ordered network. First, because symmetric direct relations can lead to asymmetric indirect paths and reachability is not necessarily transitive, matrix power approaches for reachability (Harary 1969) will not work. Geodesic search algorithms, such as Moore's (1957) BFS algorithm, would seem an efficient alternative, since every node is reached at least once. However, in many cases a pair will be reachable, but not through the geodesic. Third, there may be multiple paths from node *v* to node *w* of different lengths, only some of which may lead to another node *q*.[a] Finally, relations function differently in different paths (the same edge will be early in one path, but late in another) and thus no simple transformation of the observed graph into a directed graph usually exists.

Below, I present an efficient search procedure that involves traversing the network based on relationship timing.[b] The search algorithm enumerates all paths from a start node *s* to all nodes reachable from *s* as a tree, which can then be repeated for every node in **G**. From this information, one can identify all reachable pairs, the number of paths between nodes *s* and *t* and the length of each path. It would be trivial to modify the algorithm to record the starting and ending dates of the paths as well. The algorithm is implemented as modified breadth first search and tests the admissibility of each new arc in the time-ordered tree. To ensure that the program converges, one must forbid cycles in the paths (ego cannot infect himself or herself), and thus the longest possible path in a network is G – 1 (a path through all alters in the network).[c]

The graph is stored as an adjacency list, with a corresponding list of start times and another of finishing times. As it progresses, the algorithm constructs **P**, represented as a two-column matrix that captures the path structure from *s*. An example graph is given in Figure A1.

FIGURE A1



APPENDIX A: Identifying Indirect Paths in Time-Ordered Graphs (Cont'd)

Paths from *s* to all others are identified through a moving two-step window. For simplicity, I will refer to the first node in the two-step window as *v*, the second as *w* and the third as *q* ($v \longrightarrow w \longrightarrow q$). The algorithm used to search time-ordered paths in a graph is given in pseudo code in code A1.

Code A1: The T_BFS Algorithm to Identify Reachability in a Time-Ordered Network

```
T_BFS (G, ST, FN, s);
 STF = ST; /* Copy of the Start Dates. Will be changed to account for intersections */
 L = s and nodes adjacent to s; /* A first-in-first-out list of nodes visited by the
              algorithm */
 P = {s 1}//(alt .); /* The path search record*/
 A=G[s,]; /* A list of active relation. Initialize as all nodes in G adjacent to s */
 cnt=0; /* counter for v */
 pc=1; /* counter for P */
 Do while L is not empty;
  cnt=cnt+1; /* updates for every node in L */
  v = L(1); /* v is the 1st element of L */
  W = A(1); /* W is the first row in A */
   Do for each node w in W;
        Add an empty row to A.
        pc=pc+1; /* only updates when vw exists */
        P(pc,2)=cnt; /* update P */
              Q=set of nodes adjacent to w that are not v.
              Do for each node q in Q;
         Ancestor(cnt, q, P);
         If (STF(vw) <= finish(wq) & q is not an ancestor of w) then do;
                     add q to L;
                     add q to the last list on A. /* q is in the active list of w */
                     add {q .} to the end of P.
                     If ST(wq)<STF(vw) then do; /* an intersection */
                      If STF(wq) ≠ ST(wq); /been updated previously/;
                      STF=min(STF(wq),ST(vw));
                      End;
                      Else do;
           STF(wq)=STF(vw);
                     End;
                   End;
               End;
              End;
        End;
  Remove the first element of L;
  Remove the first row of A;
End;
Return(P);
```

APPENDIX A: Identifying Indirect Paths in Time-Ordered Graphs (Cont'd)

To start, we initialize new data structures. L is a list of nodes to visit starting with *s* and the nodes adjacent to *s* in G. L controls the progress of the algorithm. The first node in L is the start node for all two-step searches in the algorithm. New nodes, *q*, are added to the end of L every time a usable two-step path from *v* is identified. Nodes are removed from the front of L when all two-step paths from *v* have been identified. A second list A is constructed to mirror L. Each row of A contains those nodes adjacent to L_i that are active in the current path.[d] Like L, rows of A are deleted from the top when *v* is visited and added to the bottom as *q* is identified. P is constructed in the same order as L and A. Every time a new node, *q*, is found admissible it is added to the first column of P. The row in P where *w* was added is recorded in the second column.

The key to the two-step search is that inadmissible paths are those with an "early-late-earlier" pattern. Infection can only flow forward in time, and thus every two-step path through the graph must be along edges marked 'early-later.' The two-step search ensures that no 'late-earlier' steps are added to the path graph. To proceed, we initialize L, P, A, and two counters, *lc* and *pc*, that track our progress in the graph. Counter *lc* is updated every time a new node is pulled from L, while *pc* is updated only when an arc from *v* to *w* is found. Let *v* be the first element in L, and update *lc*. Next, identify all nodes adjacent to *v* on the active list; call the set W. For each node *w* in W, we update *pc* and P. By the construction process, *v* is the parent of *w* and *lc* tracks *v*. Thus the parent of *w* – p(*w*) – must equal *lc*. Because *pc* is updated only when W is not empty, *pc* tracks the children of *v*. We update P by placing *lc* in the second column of the *pc* row of P. We next add an empty row to the end of the active adjacency list, A. This will be accessed later when *w* becomes *v*. Now, select all nodes Q adjacent to *w* on G that are not *v*.

For each node *q* in Q, we evaluate if the *wq* arc is usable in the path from *vw*. An arc is usable if *q* is not an ancestor of *v* and *wq* follows *vw* in time. If *q* is not an ancestor of *v* and if the starting date of *vw* is less than or equal to the finishing date of *wq*, then the arc is usable and we add *q* to L, *q* to the active list for *w* at the end of A, and *q* to the last row, first column of P. In order not to pass through an intersecting relationship, we must push the original starting date through the network. If *wq* started before *vw* (*vw* intersects *wq*), then we change the starting date for *wq* to that of *vw*. This ensures that later in the algorithm, when *wq* is the first edge in the two-step series, we do not inadvertently go backward in time. Because *wq* can be reached from many alternate *vw* paths, arcs can multiply intersect. To ensure that all potential paths are found, we need to push the earliest intersection date through the network. If we have previously updated the starting date of *wq*, we use the earlier of the two dates (min(stf[*wq]*, st[*vw]*)). Once all elements of W have been searched, remove *v* from L and the first row of A. When L is empty, all possible paths have been found.

APPENDIX A: Identifying Indirect Paths in Time-Ordered Graphs (Cont'd)

P, the output of the algorithm, is a linked list. The first column of **P** is the node name and the second column gives the place on **P** where the parent of the node ($\mathbf{P}_i$) can be found. Consider an example from Figure A1. A time-ordered walk through G from node 3 would proceed as represented in the directed tree T. For example, there is a path 3 → 4 → 6 → 5. This path is represented in **P** by walking up **P** from node 5 at the bottom. The last row of **P** says node 5's parent is in row 6 of **P**. Moving up to row 6, we see that the parent of 5 is node 6, and that the parent of 6 is in row 3 of **P**. **P**(3,1) is 4, and the parent of 4 is in row 1. **P**(1,1) is 3, and thus we are at the end of the path. All nodes reachable from s are listed in the first column of P. The number of times a node is listed in the first column of **P** equals the number of paths from *s* to the node, and the distance(s) from *s* to any given node can be traced as in the path from 3 to 5 above.

DETECTING CYCLES IN THE PATH SEARCH

Time_BFS is a two-step search: *v* → *w* → *q*. If the time order is correct and *w* →*q* would not create a cycle, *w* →*q* is added to the path graph. To ensure that a new edge added to the path graph does not create a cycle, one must keep track of each ancestor of *wq*, which is contained in the structure of **P**. A simple algorithm that walks through **P** can thus identify whether *q* precedes *v* on the current path. If, while walking up **P** from *v*, we encounter *q*, then *q* must be an ancestor of *w*. An example of such an algorithm is given below in code A2. The function determines if *q* is an ancestor of *v*. This algorithm takes three arguments: the row in **P** where *v* is found, *q*, and **P**, and returns a 1 if *q* is an ancestor of *v*, 0 otherwise.

Code A2: Ancestor Loop for Identifying Cycles

```
Ancestor(rp, q, P);
 st_node=rp;
 stp=0;
 do while(stp=0);
   ancest=P[stnode,2];
    if(stnode=1) then do;
      stp=1;
      ancestor=0;
    end;
    else if list[ancest,1]=alt then do;
      stp=1;
      ancestor=1;
    end;
    else do;
      stnode=ancest;
 end;
end;
```

[a] That is, there may be a path $w \rightarrow x \rightarrow z \rightarrow v \rightarrow q$, and a path $w \rightarrow v \rightarrow q$. Thus the short path from w to v does not continue to q, while a longer path from w to v can pass through v to q.

[b] *Efficient* here is a relative term, and it has been shown that identifying all paths through which infection can flow is an NP-Hard problem (see Helander & Batta 1994). I have used this program for simulation work on networks with thousands of nodes, and it performs quite quickly.

[c] While the program must forbid cycles to converge, the places where a cycle would form can be recorded, allowing one to identify all connectivity sets in the graph.

[d] Nodes q_1, q_2, and q_3 might all be adjacent to w. However, there may only be a usable path from v to w to q_1 and q_3, thus for w, q_1 and q_3 are active in the current path.