On Network Theory

Stephen P. Borgatti, Daniel S. Halgin
LINKS Center for Social Network Analysis, Gatton College of Business and Economics, University of Kentucky, Lexington, Kentucky 40508 {sborgatti@uky.edu, danhalgin@uky.edu}

Research on social networks has grown considerably in the last decade. However, there is a certain amount of confusion about network theory—for example, what it is, what is distinctive about it, and how to generate new theory. This paper attempts to remedy the situation by clarifying the fundamental concepts of the field (such as the network) and characterizing how network reasoning works. We start by considering the definition of network, noting some confusion caused by two different perspectives, which we refer to as realist and nominalist. We then analyze two well-known network theories, Granovetter’s strength of weak ties theory [Granovetter, M. S. 1973. The strength of weak ties. Amer. J. Sociol. 78(6) 1360–1380] and Burt’s structural holes theory [Burt, R. S. 1992. Structural Holes: The Social Structure of Competition. Havard University Press, Cambridge, MA], to identify characteristic elements of network theorizing. We argue that both theories share an underlying theoretical model, which we label the network flow model, from which we derive additional implications. We also discuss network phenomena that do not appear to fit the flow model and discuss the possibility of a second fundamental model, which we call the bond model. We close with a discussion of the merits of model-based network theorizing for facilitating the generation of new theory, as well as a discussion of endogeneity in network theorizing.

Key words: theory; social network; flow model; bond model; endogeneity; structure

History: Published online in Articles in Advance.

Introduction

Social network analysis (SNA) is currently popular. As shown in Figure 1, publications referencing “social networks” have been increasing exponentially over time. The interest in networks spans all of the social sciences and is rising even faster in physics, epidemiology, and biology. In management research, social networks have been used to understand job performance (Sparrowe et al. 2001), turnover (Kilduff and Krackhardt 1994; Krackhardt and Porter 1985, 1986), promotion (Burt 1992), innovation (Obstfeld 2005), creativity (Burt 2004), and unethical behavior (Brass et al. 1998). Moreover, in management consulting, network analyses are fast becoming standard diagnostic and prescriptive tools (e.g., Anklam 2007, Baker 2000, Bonabeau and Krebs 2002, Cross et al. 2000).

Despite this popularity (and, perhaps, in part because of it), there exists considerable confusion about network theorizing. Even though certain network theories are extremely well known—Granovetter’s (1973) strength of weak ties theory has been cited more than 14,000 times1—it is not unusual to read that network analysis contains no theory of its own (Salancik 1995). In this view, SNA is “just” a methodology, and what theory there is “belongs to” other fields, such as social psychology. Moreover, as the term “social network” gains caché, it is increasingly applied to everything from a trade association to a listserv to a social media website such as Facebook.

Our objectives in this paper are to clarify the concept of social network and to begin to identify the characteristic elements of social network theorizing. We have a particular interest in explicating the mechanisms used in network theory to facilitate the generation of new theory. In characterizing network theory, it is important to emphasize that our objective is not to define what should and should not be network theory. We do elaborate a view of what constitutes the heart of network theorizing, but it is worth remembering that the network analysis research program (in the sense of Lakatos 1980) is a social enterprise that includes all kinds of different researchers with different aims and backgrounds. There is a great deal of work that is part of the broader SNA research program that does not include the canonical elements we describe or that includes additional elements that are not distinctive to the field.

It should also be noted that SNA theorizing encompasses two (analytically) distinct domains, which we refer to as “network theory” proper and “theory of networks.” Network theory refers to the mechanisms and processes that interact with network structures to yield certain outcomes for individuals and groups. In the terminology of Brass (2002), network theory is about the consequences of network variables, such as having many ties or being centrally located. In contrast, theory of networks refers to the processes that determine why networks have the structures they do—the antecedents of network properties, in Brass’s terms. This includes models of who forms what kind of tie with whom, who becomes central, and what characteristics (e.g., centralization or small-worldness) the network as a whole will have. In this paper, we focus on network theory proper,
although we do find it useful to make a few comments about theory of networks as well. In addition, we devote a section of this paper to assessing whether considering network theory without simultaneously treating theory of networks harms the understanding of either.

What Is a Network?

A network consists of a set of actors or nodes along with a set of ties of a specified type (such as friendship) that link them. The ties interconnect through shared end points to form paths that indirectly link nodes that are not directly tied. The pattern of ties in a network yields a particular structure, and nodes occupy positions within this structure. Much of the theoretical wealth of network analysis consists of characterizing network structures (e.g., small-worldness) and node positions (e.g., centrality) and relating these to group and node outcomes.

It is important to realize that it is the researcher—by choosing a set of nodes and a type of tie—that defines a network. To appreciate the point, consider the boundary specification problem (Laumann et al. 1983), which refers to the question of how to select which nodes to study. The naïve concern is that we may select nodes “incorrectly,” accidentally excluding nodes that should have been there and possibly including nodes that should not have been. In reality, however, the choice of nodes should not generally be regarded as an empirical question. Rather, it should be dictated by the research question and one’s explanatory theory. For example, we may be interested in how centrality in an organizational communication network is related to work performance. Therefore, we study all communication ties among all members of the organization. In making this choice, no claim is made that only ties with other members of the organization exist or matter, but rather that position in the network defined by this kind of tie among this set of actors has a measurable effect on performance. A different researcher might be interested in how a person’s communications outside the organization interact with the intraorganizational communication network to affect performance. Yet another researcher, perhaps a psychologist, might ignore the influence of others altogether (whether inside or outside the organization) and focus on how personality or life experiences affect a person’s performance.

In our view, part of the angst involved in the boundary specification problem is due to confusing networks with “groups.” A fundamental part of the concept of group is the existence of boundaries. Even though we recognize that boundaries may be fuzzy or uncertain (e.g., there are part-time members, wannabees, conflicting views of what the group is, etc.), the distinction between insiders and outsiders is an important part of the group concept. Therefore, when studying groups, we are justifiably concerned with establishing the boundaries of the group. For example, if we are studying gangs in Los Angeles, we would not want to approach the boundary specification problem in a wholly etic way, such as defining gang members as all young males living in a given area.

In contrast to groups, networks do not have “natural” boundaries (although, of course, we are free to study natural groups, in which case the group boundaries determine our nodes). Networks also do not have to be connected. A disconnected network is one in which some nodes cannot reach certain others by any path, meaning that the network is divided into fragments known as components (see Figure 2). For those confusing networks with groups, this may seem an odd conceptualization. The advantage, however, is that it facilitates the study of network evolution. For example, suppose we study the freshman class at a university, focusing on friendships. Initially, it may be that none of the freshmen is friends with any other, defining a maximally disconnected network with as many components as nodes. Over time, friendships begin to develop, and the number of components may reduce rapidly. Eventually, it is possible that all the actors are connected in a single component in which every node can be reached from every other by at least one path (even if very long). Thus, by allowing the network to be disconnected, we can trace the evolution of connectivity within it. In this perspective, we do not ask “under what circumstances will networks emerge” (Kahler 2009, p. 104), as if they were groups. Rather, we ask how specific properties of
the network, such as the level of fragmentation or characteristic path length, change over time.

A closely related issue is what “counts” as a tie. A common beginner’s question is, which network questions should I ask in order to get at the network? Implicit in the question is the idea—labeled the realist position by Laumann et al. (1983)—that there is a “true” network of relationships out there, and our job as researchers is to discover it. Given that assumption, it is reasonable to ask which social network questions have proven effective at eliciting this network. However, a more sophisticated view of social networks—labeled the nominalist position by Laumann et al. (1983)—holds that every network question (such as “Who are you friends with?” or “Who do you seek advice from?”) generates its own network, and which to use is determined by the research question. Thus, a given research question may lead us to examine the advice and friendship ties within an organization, whereas another research question may lead us to examine “who-likes-whom” ties. No matter what kind of tie we are interested in, measuring that kind of tie among all pairs of nodes in the sample defines a network, and each network will have its own structure and its own implications for the nodes involved. For example, being central in a gossip network might be entertaining and beneficial, whereas being central in a who-dislikes-whom network might be painful and deleterious.

In practice, the kinds of ties that network theorists tend to focus on can be categorized into two basic types: states and events (see Table 1). States have continuity over time. This is not to say they are permanent, but rather that they have an open-ended persistence. Examples of state-type ties include kinship ties (e.g., parent of), other role-based relations (e.g., friend of or boss of), cognitive/perceptual relations (e.g., recognizes or knows the skills of), and affective relations (e.g., likes or hates). State-type ties can be dimensionalized in terms of strength, intensity, and duration.

In contrast, an event-type tie has a discrete and transitory nature and can be counted over periods of time. Examples of event-type ties include e-mail exchanges, phone conversations, and transactions such as sales or treaties signed. Cumulated over time, event-type ties can be dimensionalized in terms of frequency of occurrence (e.g., the number of e-mails exchanged). It is these kinds of ties that researchers have in mind when they define networks as a “recurring pattern of ties” (e.g., Dubini and Aldrich 1991, Ebers 1997).

Both state-type ties and event-type ties can be seen as roads or pipes that enable (and constrain) some kind of flow between nodes. Flows are what actually pass between nodes as they interact, such as ideas or goods. Hence two friends (state-type social relation) may talk (event-type interaction) and, in so doing, exchange some news (flow). As we discuss in the next section, one large swath of network theory is about how position in a buckcloth network determines the timing or quantity of flows to the actor occupying that position.

We might also note that, in empirical studies, researchers often make use of relational states and events that are not, properly speaking, social ties. For example, a frequent proxy for social ties is group comembership, such as being on the same board of directors or belonging to the same club. Similarly, coparticipation in events such as parties is used as a proxy for unobserved social relationships. Other dyadic variables of this type include geographic proximity (Allen 1977) and similarity of traits such as behavior, beliefs, and attitudes (McPherson and Smith-Lovin 1987, McPherson et al. 2001). From a theoretical point of view, comemberships, coparticipations, geographic proximities, and trait similarities can all be seen either as dyadic factors contributing to the formation of ties (e.g., meeting the other members of your club) or as the visible outcomes of social ties (as when close friends join the same groups or spouses come to hold similar views).

Network Theorizing

To illustrate the nature and distinctive flavor of network theorizing, we start by describing in detail two well-known network theories, and we then analyze them for their key characteristics. We begin with Granovetter’s (1973) strength of weak ties (SWT) theory and then move to Burt’s (1992) structural holes (SH) theory.

The SWT theory is organized as a set of explicit premises and conclusions. The first premise of the theory is that the stronger the tie between two people, the more likely their social worlds will overlap—that they will have ties with the same third parties. As a result, if A and B have a strong tie, and B and C have a strong tie, the claim is that A and C have an increased chance of having at least a weak tie (e.g., A and C are acquaintances). This is a kind of transitivity—one that some authors have called $g$-transitivity (Freeman 1979).

The reason for this transitivity, Granovetter argues, is that the underlying causes of tie formation have this kind of transitivity built into them. For example, people tend to be homophilous, meaning that they tend to have stronger ties with people who are similar to themselves (Lazarsfeld and Merton 1954, McPherson et al. 2001).

<table>
<thead>
<tr>
<th>State-type ties</th>
<th>Event-type ties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinship ties (e.g., brother of)</td>
<td>Interactions (e.g., giving advice to; sending e-mail)</td>
</tr>
<tr>
<td>Other role-based ties (e.g., boss of or friend of)</td>
<td>Transactions (e.g., signing treaty with; making a sale)</td>
</tr>
<tr>
<td>Cognitive (e.g., knows)</td>
<td></td>
</tr>
<tr>
<td>Affective (e.g., likes or dislikes)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Types of Social Ties
Homophily is weakly transitive because if A is similar to B, and B is similar to C, then A and C are likely to be somewhat (i.e., weakly) similar as well. To the extent that similarity causes ties, this will induce weak transitivity in the tie structure as well.

The second premise of SWT is that bridging ties are a potential source of novel ideas. A bridging tie is a tie that links a person to someone who is not connected to his or her other friends. The idea is that, through a bridging tie, a person can hear things that are not already circulating among his close friends. In Figure 3, A’s tie with G is a bridging tie. Because A is the only person in her social group with a tie outside the group, A has the benefit of hearing things from G that the rest of A's group has not yet heard.

Putting the two premises together, Granovetter reasons that strong ties are unlikely to be the sources of novel information. The reason is as follows. First, bridging ties are unlikely to be strong. According to the first premise, if A and G have a strong tie, then G should have at least a weak tie to A’s other strong friends, which would imply that the A–G tie was not a bridge, because there would be multiple short paths from A to G via their common acquaintances. Therefore, it is only weak ties that are likely to be bridges. Second, because bridges are the sources of novel information, and only weak ties are bridges, it is the weak ties that are the best potential sources of novel information. Granovetter uses this theory to explain why people often get or at least hear about jobs through acquaintances rather than close friends. In this sense, the theory is one of individual social capital, where people with more weak ties (i.e., more social capital) are more successful. Granovetter also applies the theory at the group level, arguing that communities with many strong ties have pockets of strong local cohesion but weak global cohesion, whereas communities with many weak ties have weak local cohesion but strong global cohesion. Using the case study of Boston, in which the city assimilated one adjacent community but failed to assimilate another, he suggests that a community’s diffuse, weak-tie structure constitutes group-level social capital that enables the group to work together to achieve goals, such as mobilizing resources and organizing community action to respond to an outside threat.

Another well-known network theory is Burt’s (1992) structural holes theory of social capital. The theory of SH is concerned with ego networks—the cloud of nodes surrounding a given node, along with all the ties among them. Burt argues that if we compare nodes A and B in Figure 3, the shape of A’s ego network is likely to afford A more novel information than B’s ego network does for B, and as a result, A may perform better in a given setting, such as an employee in a firm. Both have the same number of ties, and we can stipulate that their ties are of the same strength. However, because B’s contacts are connected with each other, the information B gets from, say, X may well be the same information B gets from Y. In contrast, A’s ties connect to three different pools of information (represented by the circles in Figure 4). Burt argues that, as a result, A is likely to receive more nonredundant information at any given time than B, which in turn can provide A with the capability of performing better or being perceived as the source of new ideas.

Kilduff (2010) argues that Burt’s portrayal of the social world differs significantly from that of Granovetter along a variety of dimensions. For example, Kilduff sees Granovetter as embracing a serendipitous world in which people form ties that only incidentally prove useful, whereas Burt embraces a more strategic and instrumental view. However, at the level of the specific theories of SWT and SH, it should be obvious that Burt’s theory is closely related to Granovetter’s. In Burt’s language, A has more structural holes than B, which means A has more nonredundant ties. In Granovetter’s language, A has more bridges than B. But whether we call them nonredundant ties or bridges, the concept is the same and so are the consequences: more novel information. Where Granovetter and Burt differ is that Granovetter further argues that a tie’s strength determines whether it will serve as a bridge. Burt does not disagree and even provides empirical evidence that bridging ties are weaker in that they are more subject to decay (Burt 1992, 2002). However, Burt sees tie
strength as a mere “correlate” of the underlying principle, which is nonredundancy (1992, p. 27). Thus, the difference is between preferring the distal cause (strength of ties), as Granovetter does, and the proximal cause (bridging ties), as Burt does. The first yields an appealingly ironic and counterintuitive story line; the second “captures the causal agent directly and thus provides a stronger foundation for theory” (Burt 1992, p. 28). In addition, Granovetter uses getting jobs as an outcome of having nonredundant information, whereas Burt uses getting promoted. In our view, these are small differences in ornamentation. Both theories are based on the same underlying model of how networks work.

Characterizing Network Theory
Examing SWT and SH from a metatheoretical point of view, we can see two features of network theory that are highly characteristic. First, the twin notions of structure and position play fundamental roles. For example, in SWT, the reason weak ties are useful is not because they are inherently so but because it is the weak ties that tend to bridge network clusters. It is their structural role that makes them advantageous. Similarly, in SH, it is the shape of the ego network around a person that confers advantages to that person. Note that the theory ignores egos’ own attributes (such as how creative they are) and also the attributes of egos’ contacts (e.g., how smart they are, or how gullible or powerful) and only looks to see whether the alters are numerous and unconnected. This is not to say that ego and alter attributes are not important (they may well be much more important); it is just that the agenda of the theory—and the charter of network theory in general—is to explicate the connection between structure and outcome, and one aspect of this agenda is the study of the pure effects of structure. To be clear, the general agenda of examining the consequences of network structure includes the examination of how structure and attributes interact to yield outcomes. But a piece of that investigation is the exploration of how structural differences alone have effects.

Second, there is an implicit theory of network function; in the case of SWT and SH (but not all network theories), the network function is the flow or distribution of information. In effect, SWT and SH rely on an underlying model of a social system as a network of paths that act as conduits for information to flow. We refer to this as the flow or pipes model. The abstract flow model carries with it some basic assumptions, such as the longer a path is, the longer it takes something to traverse it. From this general model, we can readily derive a number of theoretical propositions that form the core of theories like SWT and SH. For instance, nodes that are far from all others will, on average, receive flows later than nodes that are more centrally positioned. Similarly, nodes that are embedded in locally dense parts of a network will often receive the same bits of flow from their various contacts, because the contacts are tied to each other as well. These flow outcomes (time until arrival; amount of nonredundant flow received) are then related to a variety of more general outcomes, such as creativity, likelihood of promotion, getting a job, etc.

As an aside, by specifying additional features of the central process in the model, we can generate additional theoretical implications. For example, for the flow model, we can specify different variations for how flows move through the network (Borgatti 2005). For instance, a dollar bill moving through a network transfers from actor to actor in such a way that it is never in two places at the same time. In contrast, a virus or bit of news duplicates from actor to actor so that when A passes it to B, A retains a copy. Another dimension of how things can flow is what kinds of paths or trajectories they trace through the network. Some viruses, for instance, tend not to reinfect a node, either because the node has become immune or because it is now isolated or dead. In network theory, this kind of trajectory is known as a true path. In contrast, gossip can easily pass through a node multiple times, because transmitters do not necessarily know who has already received it. However, we might argue that gossip does not revisit ties. That is, once I have told you a story, I am unlikely to tell you the same story again (forgetfulness just adds a bit of noise to the system). In network theory, this kind of traversal is known as a trail. A dollar bill illustrates yet another type of path, technically called a walk, which is unrestricted with respect to whether it reuses nodes or ties (e.g., on a given day, a person might give the bill to a store and the next day receive it back in change for another purchase. It could then be given by that person to that store yet again in a third transaction).

Given that things flow through the network according to certain rules, some obvious outcomes can be predicted as consequences of the network structure. For instance, at the node level, we may be interested in the expected time until (first) arrival of whatever is flowing through the network. Certain (central) nodes are positioned in such a way that, on average, they receive the flow sooner than other nodes. We may also be interested in how often or with what level of certainty a node receives a given bit of flow. It should be noted that both of these flow outcomes are fully defined within the model but are not necessarily the empirical outcomes that we actually measure. Network research consists of equating these model outcomes with other constructs such as the likelihood of getting a job (Granovetter 1973, 1974), being promoted (Brass 1984, 1985; Burt 1992), or being creative (Burt 2004, Perry-Smith 2006). Thus, hypotheses that are actually tested in empirical studies relate features of the observed network to outcomes such as performance in an organizational setting, and network theory consists of elaborating how a given network structure interacts
with a given process (such as information flow) to generate outcomes for the nodes or the network as a whole.

The flow model is the most developed theoretical platform in network theory, but it is not the only one. The field has clearly identified phenomena and developed theoretical explanations that cannot be reduced to the flow model. One such area is the study of power. Cook and Emerson (1978) pioneered the experimental study of the exercise of power in exchange networks. In their experiments, subjects occupied nodes in a network designed by the researcher. The subjects played a game in which, at each round, they had the option to negotiate a deal with someone they were connected to. At each round, each subject could only close on one deal. Across rounds, the subjects’ objective was to make as many deals at the best possible terms as possible. For example, for the network in Figure 5, Cook and Emerson found that the subject in position B was able to negotiate the best deals, even though subjects were not shown the structure of the network they were embedded in. From Cook and Emerson’s point of view, the fundamental advantage that B enjoys is the dependency of others, which is a function of the (lack of) availability of alternatives on the part of B’s potential partners. Node B has two alternatives available for making a deal, whereas A and C have no alternatives to B and are therefore wholly dependent on B. This positional advantage is very different from the concept of centrality, which largely emerges from the flow model. This can be seen in the experimental results for the network in Figure 6, in which B and D emerge as high-power positions and A, C, and E have very low power. This might seem surprising given that B, C, and D all have two potential trading partners, but the difference is that C’s partners B and D both have better alternatives to C, namely, the wholly dependent A and E. Thus, whereas a basic principle in centrality phenomena is that being connected to well-connected others implies greater centrality, in power phenomena it can be the other way around: being connected to weak others makes one powerful, and being connected to powerful others makes one weak (Bonacich 1987, Markovsky et al. 1988, Marsden 1983).

What is especially interesting about network power is that network structure (and location within that structure) matters, and yet the basis for network power is not the accumulation or early reception of a resource that is flowing over well-positioned nodes, as it is in the flow model. This is especially clear in the experimental setting because the rules of the game explicitly prohibit the flow of resources. Nor is power itself flowing, because if it were, nodes adjacent to a powerful node would be empowered.

Another way to look at network power is in terms of coordination and virtual amalgamation. Consider node E negotiating with a set of alters A1–A4, as shown in Figure 7. Because there is only one E and several As, one might expect E to have a difficult time; however, this would only be true if the As worked together as a unit. One way this can happen is if the As are bound together by ties of solidarity. In the extreme, this can be seen as converting the As into a single node that can deal with E on an equal basis—that is, a mechanism of virtual amalgamation, as shown in Figure 8. This is the principle behind unionization.
A closer look at the unionization example suggests that two kinds of relationships among nodes are implicit. One is the potential exchange tie that exists between E and its alters. The other is the solidarity ties that may exist among the alters. A key point is that the alters in Figures 7 and 8 are essentially of a type, with the same interests (e.g., to exchange with E) and capabilities (making them interchangeable from E’s point of view). This effectively allows E to induce competition (itself a kind of tie) between the As (which can be mitigated by ties of solidarity, as in unionization). A different sort of case is the so-called “network organization,” in which a set of autonomous organizations coordinate closely, as if comprising a single superordinate entity (Jones et al. 1997, Powell 1990). By working together they can accomplish more than they could alone. We can view this phenomenon as virtual capitalization, meaning that the bonds between the nodes enable the nodes to act as if they were transferring the capabilities of the other nodes to each other, but without actually doing so.

Supply chain networks have a similar character. Rather than vertically integrating and taking on the tasks and abilities of upstream suppliers (as in the simple capitalization process that occurs in the flow model), the firm has bonds with those suppliers that enable it to behave as if it had those capabilities. More generally, this is the same phenomenon studied by principal/agent theory, in which the agent acts in the interests of a principal without the principal having to do the work itself.

In all these examples—from exchange experiments to principals and agents—a common underlying theme is that the network tie serves as a bond that aligns and coordinates action, enabling groups of nodes to act as a single node, often with greater capabilities. The bonding function serves as the basis for what we call the bond or coordination model, and it is the analogue of the flow function in the flow model.

An interesting question is whether the work on experimental exchange networks can be derived from the bond model. There is a point of commonality, which is that when a pair of nodes makes a deal in a given round, the nodes become, momentarily, a unit that excludes those not part of the deal. From this perspective, a node’s advantage derives from its inexcludability. For example, consider the positions of B and D in Figure 6: if C and D make a deal, B can make a deal with A. If D makes a deal with E, then B can make a deal with either A or C. There is no combination of outcomes in any round that does not give both B and its twin D the option of making a deal.

Like the flow model, the bond model permits a number of derivations, which in turn enable us to construct measures of power and predict power-based outcomes. For example, the logic of dependency and excludability dictates that the existence of node A on the other side of node B is detrimental for node C, who would rather that B not have any alternatives to itself. Generalizing this a bit, paths of an even length emanating from a node reduce its power, whereas paths of an odd length increase its power. This theorem is the basis for several measures of network power, including the graph-theoretic power index (Markovsky et al. 1988) and beta centrality (Bonacich 1987, 2007).

Another derivation from the bond model is that isomorphic nodes will have similar outcomes even if they are not reachable from each other (as flow-based processes would require). For example, in Figure 9, nodes A and H are structurally isomorphic and therefore must have the same structural advantages and disadvantages. Holding constant individual differences in the abilities of actors occupying network positions, we can expect that structurally isomorphic nodes will have similar outcomes.

The third thing to note about network theory is that the core concept of the field—the network—is not only a sociological construct but also a mathematical object. As a result, it is sometimes possible to use the machinery of mathematics to generate new theory. For instance, Rapoport (1963) and others show that transitivity tends to create highly clustered graphs that have many long paths or disconnected components, which means that networks with high transitivity are slow or incomplete diffusers. This, of course, is the basis for SWT and SH theory. More generally, the coincidence of sociological networks and mathematical networks makes it easy to generate formal theory that is expressed in mathematical form. This can be a blessing, but it also carries with it the danger that the nonmathematically inclined will not see it as theory at all, but rather as some form of statistics. A good example is the notion of betweenness centrality, which is defined by the formula shown in Equation (1). It has been shown (Borgatti 2005) that...
the betweenness formula gives the expected values of the number of times something reaches a node in a certain flow process (namely, one in which the things flow along shortest paths, and when there are multiple equally short paths, they choose one of them with equal probability). Thus, what looks like methodology is in fact formal theory based on the flow model:

$$b_k = \sum_{i,j} g_{ikj} / g_{ij},$$

(1)

where $b_k$ is the betweenness of node $k$, $g_{ij}$ is the number of geodesic paths from $i$ to $j$, and $g_{ikj}$ is the number of geodesic paths from $i$ to $j$ that pass through $k$.

It is worth noting that even things as technical as the notions of structural equivalence (Lorrain and White 1971) and regular equivalence (Everett and Borgatti 1994, White and Reitz 1983) were explicitly developed in an effort to formalize the social role theory of Linton (1936), Nadel (1957), Merton (1959), and others. Similarly, the notions of clique (Luce and Perry 1949), $\alpha$-clique (Luce 1950), $k$-plex (Seidman and Foster 1978), and other subgroups, which sound so methodological, were actually attempts to state with mathematical precision the concept of group that Cooley (1909), Homans (1950), and others had discussed at a more intuitive level.9

Goals of Network Theorizing

So far, we have focused on characterizing modes of explanation in network models. In this section we focus on characterizing the kinds of outcomes that these models are used to explain. As in much of social science, there are two generic types of outcomes that network research has sought to explain. The first can be broadly termed choice and includes behaviors, attitudes, beliefs, and (in the case of collective actors like organizations) internal structural characteristics. Network research on choice has often been framed in terms of similarity of choice, as in explaining which pairs of nodes make similar choices. For this reason, work in this area is often referred to as the social homogeneity literature, as noted by Borgatti and Foster (2003). The second generic outcome is success, which includes performance and rewards, whether at the node or whole network level. Work in this area is known as the social capital literature. Combining these two generic outcomes with the two explanatory models we have outlined, we get a simple typology of network theorizing.

As shown in Table 2, the top right quadrant, contagion, consists of flow-based explanations of (similarity of) choice, which is a well-populated segment of the literature. The principal example of this kind of work is diffusion or adoption of innovation studies in which nodes are conceptualized as influencing each other to adopt their traits. For example, work in the organizational theory literature posits that one reason organizations have similar structures is diffusion (Davis 1991, DiMaggio and Powell 1983). Extending Dimaggio and Powell (1983), we can use locus of agency to distinguish four different types of diffusion. As shown in Table 3, DiMaggio and Powell discuss mimetic processes, in which the adopter actively seeks to copy a trait from a node in its environment, and coercive processes, in which the node is forced by a node in its environment to adopt a trait (such as a certain accounting system). In addition, we identify two other processes, the apprentice process, in which both the ego and its environment are actively trying to help the ego get what the alter has, and the osmotic process, where neither party is actively expending energy to enable the transfer, but it happens anyway (as when the ego learns a new term or concept simply by listening to the alter).

The bottom right quadrant of Table 2, convergence, contains bond-based explanations of homogeneity. Work in this area includes research on structural equivalence (Lorrain and White 1971, Burt 1976), which posits that nodes adapt to their environments, and as a result nodes with similar structural environments will demonstrate similarities (Erickson 1988). For example, suppose two people in different parts of the world are highly central in the advice networks around them; that is, everybody is constantly seeking their advice. As a result, both of them develop a dislike of the phone, because it so often brings work for them. Hence, sameness in terms of centrality level leads to similar attitudes.

Work in this area can also be seen as a special case of coordination in which nodes behave similarly rather than simply in concert (as in the bottom left quadrant), which is similar to the sociological concept of gemeinschaft (Tönnies 1912). Other work we would classify...
in this quadrant includes the networks-as-prisms concept of Podolny (2001), along with the empirical work of Kilduff and Krackhardt (1994) and the identity-based network research of Podolny and Baron (1997) and Halgin (2009), which suggests that network ties provide informational clues to audiences regarding the quality and identity of an actor.

The top left quadrant of Table 2, **capitalization**, contains flow-based explanations of achievement. The basic concept here is that social position in a network provides access to resources. Work in this area is exemplified by strength of weak ties theory (Granovetter 1973), Lin’s (1988) social resource theory, and the information benefits theory of structural holes (Burt 1992).

Finally, the bottom left quadrant of Table 2, **cooperation**, consists of bond-based explanations of achievement. Here, combinations of nodes act as a unit, excluding others and exploiting divisions among them. This is exemplified by the stream of research on experimental exchange networks (Bonacich 1987, Cook and Emerson 1978, Markovsky et al. 1988), as well as the control benefits theory of structural holes (Burt 1992). For a more detailed discussion of the work falling into each quadrant, see the review by Borgatti and Foster (2003).\footnote{10}

**Discussion**

In this section, we comment on the analysis presented in this paper, organizing our comments into two sections, entitled model-based theorizing and endogeneity.

**Model-Based Theorizing**

In this paper, we have argued that at least some portions of network analysis can be described as model-based theorizing and have outlined two fundamental models, the flow and bond models, that underlie extant network theorizing. According to Lave and March (1975), model-based theorizing is one of the strongest forms of theorizing. In model-based theorizing, we imagine an observed state of affairs as the outcome of an unseen process, which is what is specified by the model. Given the model, you can derive testable implications, including the original observations that led you to postulate the model. Ideally, a model can also be expressed formally so that the machinery of mathematics and/or simulation can be used to derive additional implications that might be difficult to develop by simple intuition (e.g., Everett and Borgatti 1994, Lorrain and White 1971, Luce and Perry 1949, White and Reitz 1983). The implications are used to test the theory as well as to apply the theory to new situations.

One feature of model-based theorizing is the separation between the abstract elements of the model and the mapping of those elements to the real world. Hence, we should write network theory at the level of, say, the function of enabling something to flow from one node to another, not at the level of, say, who-likes-whom ties. For example, in SWT, Granovetter (1973, p. 1361) specifies quite clearly what a strong tie is (namely, a combination of time, emotional intensity, intimacy, and reciprocal services). However, this definition is open to debate and is not appropriate in all settings, such as when the nodes are firms. A closer look at the theory shows that a specific definition is actually unnecessary: any type of tie that has the property of generating g-transitivity will do. The rest of the theory does not make use in any way of the fact that strong ties were defined in terms of emotional intensity and the rest. The only property of strong ties that is actually utilized is the property of g-transitivity.

There is an analogy here to object-oriented computer programming (OOP), in which real-world entities are modeled as classes of “objects” that consist of data along with procedures (called “methods”) that operate on them. A key principle of OOP is that one should program to an interface rather than to an implementation. What this means is that higher-level code should not have to know the details of how lower-level code works—the functions of the lower-level code should be encapsulated so that higher-level code deals only with the functions of the code, not the means by which they are accomplished. For example, if we are modeling interactions among animals, our main code should not have to know how, exactly, a cow sounds or a dog moves. Rather, it should be able to issue to the object representing a particular animal a general command such as “make sound” or “move” and have this interpreted appropriately by the object, which knows how to make its own sound and how to execute its own way of moving. In this way, any changes to how a specific kind of animal makes sounds or moves will not affect the main program, and new types of animals can easily be incorporated.

In network theory, the concept corresponding to OOP’s object is the network, and what corresponds to OOP’s methods is the set of processes or functions that we define on the network, such as flow of information. In our view, this analogy helps point the way toward dealing with issues of context and culture. For example, a theory built on a particular definition of tie (e.g., friendship) will run into problems when we try to apply it cross-culturally, because friendship has different implications in different cultures and settings. A better approach is to build theory at the level of abstract ties that have certain properties needed by the theory (e.g., ties create shared identity, or ties transfer resources). Then to apply the theory in a given setting, we use our situated ethnographic knowledge to find an appropriate, specific type of tie that, in that context, entails the functions needed by the theory.

The analogy also helps clarify the question of whether we can apply the same network theories to collective
and/or nonhuman actors—such as firms—as we do to actors that are individual persons (Madhavan 2010). For example, if we wish to apply SWT theory to firms, we need only ensure that the kind of ties we study have the property of g-transitivity and serve as pipes through which resources flow. We need not worry that ties among firms do not have “emotional intensity” or “intimacy” as long as there is a kind of interfirm tie that has the two properties that Granovetter’s model depends on—namely, transitivity and enabling the flow of information. Of course, it should be noted that different kinds of nodes have different capabilities, which needs to be taken account of in generating the auxiliary theorizing that links model outcomes to such outcome variables as, say, performance or creativity. For example, when an individual hears two bits of information, he has a fighting chance of integrating them, but when a firm hears two bits of information, it may be different parts of the organization that house them, and the bits may never come together in the same space to be integrated.11

**Endogeneity**

In this paper, we have separated network theory from theory of networks in an attempt to trade breadth for depth. However, reviewers of this paper have raised some concerns about this separation. First, there is the question of whether the distinction is “merely” analytical, because it might be expected that, in reality, the two kinds of processes occur together. Second, there is the concern that we cannot correctly predict outcomes of network structure if we have not taken account of how the network got there—that is, the trajectory matters. Third, there is the question of endogeneity. Endogeneity means different things in different contexts, but one sense of the term is that factors seen as causing the outcome are in some part dependent on the outcome. Finally, the issue of agency comes to mind. If actors deliberately shape the networks around them for their benefit, can it really be said that it was network structure that led to the benefit?

To begin our discussion, let us make clear on a semantic level that network theory and theory of networks are not disjoint sets. Recall that we defined the domain of network theory as the consequences of network processes and structures. In our examples, these consequences were things such as performance or reward. However, it is obvious that the consequences of network processes can include other network phenomena, in which case network theory is simultaneously theory of networks, which is to say we have a network theory of networks (see Table 4). In a network theory of networks, both independent and dependent variables involve network properties. An example is the cascade of effects that can be produced by the formation of a positive or negative tie between two actors. For instance, suppose spouses Bill and Nancy develop a negative tie between them, culminating in an acrimonious divorce. According to balance theory, we can expect that a third person, Sally, with a strong positive tie to both parties, will experience stress and be likely to weaken the tie with one of them—i.e., choose sides. This in turn has a ripple effect on Sally’s friends, who may also be forced to choose sides. Another example is the interaction between homophily and centrality (Ibarra 1992). If actors have a marked tendency to be homophilous with respect to race, and one race has a clear numerical majority, we can expect that members of the majority race will be more central.

A more interesting question is whether, as Salancik (1995) seems to feel, a network theory must include a theory of networks. In particular, are there any circumstances where we must take into account how a network reached a given structure in order to understand the consequences of that structure? Common sense would suggest that the answer is “yes.” For example, consider two nodes who occupy identical positions in a network (both have many structural holes) and have similar motivations. However, node A reached that position through a long campaign of strategic relationship building, whereas node B arrived at it serendipitously and in fact is unaware of the potentialities of its position. We can readily imagine that, in a population of nodes like A, the correlation between structural holes and power will be higher than in a population of nodes like B, who do not think to exploit their position. Thus, the causal link between holes and power varies depending on how the nodes got their holes. Or does it? The key difference between A and B is that A’s journey to that position implies awareness of its value, allowing A to exploit it. But suppose there other ways of becoming aware of the value of one’s structural holes. For example, suppose node B attends an executive education class on social network analysis. Given that B has the same position and same awareness as A, shouldn’t the consequences for B be the same as the consequences for A, all else being equal?

Thus, on closer inspection, the answer to whether network theory must include theory of networks would seem to be “no.” If a model has been constructed that embodies the mechanisms that convert a given set of inputs at time $T$ to an output at $T + 1$, then given that input, nothing else is needed to explain the outcome. In

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Network Theory and Theory of Networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent variable</td>
<td>Nonnetwork variable</td>
</tr>
<tr>
<td>as antecedent theory</td>
<td>Theory of networks</td>
</tr>
<tr>
<td>Nonnetwork variable as antecedent</td>
<td>Theory</td>
</tr>
<tr>
<td>Network variable as antecedent</td>
<td>Network theory of networks</td>
</tr>
</tbody>
</table>
practice, however, it is a little more complicated. For one thing, knowing the input at \( T \) may involve a longitudinal analysis. As a very simplistic example, suppose an outcome is a function of whether a network is increasing or decreasing in density (e.g., the nodes make certain choices when they perceive the density to be changing in a particular direction). A snapshot of the network at a single point in time does not tell us whether the density is waxing or waning. However, once we have determined, via longitudinal analysis, whether it is waxing or waning at time \( T \), we can set the “momentum” variable at \( T \) to the observed value, and we then have all the information we need to understand what happens next.

As a more substantive example, consider the strength of weak ties theory. An appealing feature of the theory is that it spans both the theory of networks domain and the network theory domain. As discussed, a key premise of SWT is that networks form in such a way that they exhibit g-transitivity. It can then be derived that bridging ties are unlikely to be strong ties. This is the “theory of networks” portion of the theory. If we then combine another premise (that bridging ties are the most likely source of novel information), we can conclude that the structural property of having many weak ties is likely to be associated with access to more novel information, which in turn may be associated with performance gains. This is the “network theory” portion. The combination of the two portions is both satisfying and elegant. But do we need the first part to get the second part right? Strictly speaking, the answer is “no.” To derive the hypothesis that weak ties will be associated with strong performance and that this is mediated by access to novel information, we do not need to know why networks have g-transitivity—merely that they do.

However, there is also the matter of how satisfying a theory feels. In any theory in which X leads to Y, we can wonder what leads to X. In some cases this feels like an urgent and necessary question. For example, if the explanation for why people divorce is “because they want to,” we are likely to demand an explanation of why they want to. In other cases, there is enough of a sense of process or mechanism in the theory that we are willing to back off of the chain of infinite regress. For example, among other arguments, Granovetter (1973) uses balance theory to explain g-transitivity. According to balance theory, a person seeks to be congruent with those she likes. When she is not, she feels dissonance and seeks to reduce it. We could ask why, but most of us in the management field are willing to let that one go and let the psychologists deal with it. Ultimately, at what point we feel enough explanation has been given to be satisfying in a given context is a question for the sociology of science and not a question about a particular field such as network theory.

It should be noted that, in principle, the ability to theorize about consequences of networks independently of antecedents does not absolve the field from resolving issues of endogeneity in a given empirical inquiry. For example, Lee (2010) finds that in a biotech setting, the cross-sectional correlation between structural holes and innovative performance disappears when controlling for inventors’ past performance. Thus, in that particular case, it appears that it is performance that creates holes rather than the other way around, and whatever is responsible for performance is stable over time so that past performance predicts future performance. Thus, it could be an individual characteristic such as skill or personality that causes both structural holes and performance. This is an important result but should not be misread as saying something fundamental about network theorizing. In every field study we must be concerned about whether A causes B, or the other way around, or whether both are caused by an uncontrolled third variable.}

Finally, we take up the issue of agency as it relates to endogeneity. One of the legacies of the social capital approach in social network research is the notion that ties and position can be “good,” that is, associated with positive outcomes such as performance or reward. Inevitably, this leads to the following bit of reasoning: if occupying a certain position in the network is rewarding, we can expect actors to take steps to achieve that position. Thus, the network structure is not a given in the sense of an exogenous variable, but rather it is shaped by the actors specifically to achieve the very outcomes that we researchers associate with those structures. Therefore, any theory of social networks must take into account actors’ agency in creating those networks. The problem with this, as we have pointed out, is that it is not the actors’ intentions and actions leading to occupying a certain position that creates the outcome but the actual occupation of the position. A rock dropped from the same place in the same way has the same outcomes regardless of whether it was dropped on purpose or by accident. Given the same conditions, the outcomes are the same.

One thing this discussion highlights is the importance of node attributes and contextual factors in network research. Occupying a certain structural position carries certain potentialities, but the actual outcomes may depend on a number of additional factors, including how the actor plays it. How they play it may be a function of how they got there, and so knowing how they got there could give our predictive ability a boost. However, it is not the journey itself that is the theoretical variable but rather the complex of conditions (e.g., state of mind, skills, motivations) at the end of the journey that is the causal agent. If we can measure that condition directly, there is no need to code the journey. In this sense, if we find that we cannot predict how X leads to Y without knowing how X came about, it is evidence that our theory of how X leads to Y is incomplete: we are missing
a node attribute or other contextual factor that interacts with network position to bring about the outcome being modeled.

Conclusion

Our principal goal in this paper has been to dissect and characterize network theorizing. In doing so, we have argued that much of network theory (and methodology) is based on the flow model, which is now well elaborated and serves to unify large portions of network theory. We have also argued that another model, the bond model, is under development and can potentially unify several other areas of inquiry. More generally, we hope that our discussion of network concepts and model-based theorizing in network research will help clarify existing theory as well as facilitate the generation of new theory.

Acknowledgments

The authors are grateful to the Organization Science editorial team and the members of the LINKS Center for Social Network Analysis at the University of Kentucky, especially Dan Brass, Ginny Kidwell, Joe Labianca, and Ajay Mehra, for their help in shaping the authors’ thinking. This work was funded in part by Grant HDTRA1-08-1-0002-P00002 from the Defense Threat Reduction Agency.

Endnotes

1Source. Google Scholar.

2Etic versus emic is a distinction made in cognitive anthropology between organizing the world using researcher-driven criteria (etic) and organizing things the way natives do (emic). The terms come from the linguistic distinction between phonetic (how things sound) and phonemic (what things mean).

3It should be acknowledged, however, that there is a literature that labels organizational forms intermediate between hierarchies and markets as “networks.” In this literature, a network refers to a group of organizations working closely together, almost as if they were one superordinate organization.

4This is Atkin’s (1972) “backcloth/traffic” distinction.

5Granovetter (1973, p. 1361) provides a definition of strength of tie, but it is useful to realize that any definition of tie strength that preserves the first premise can be used (Freeman 1979).

6More technically, a bridge is a tie between A and B that, if removed, would leave a very long path (if any at all) connecting A to B. A bridge, then, is a shortcut in the network.

7Note that there is no claim that all weak ties are sources of novel information—just the ones that happen to be bridges. Granovetter’s point is simply that it is weak ties rather than strong ties that are more likely to be bridges.

8We are grateful to an anonymous reviewer for pointing this out.

9An explanation of these terms is beyond the scope of this paper. For a review, consult Wasserman and Faust (1994).

10The terminology in Borgatti and Foster (2003) is somewhat different, but the underlying ideas are the same.

11We are grateful to an anonymous reviewer for making this point.

12Note that the possibility that A and B both cause each other should not concern us: if our theorizing suggests that A causes B, and we find that A and B cause each other, then our theory is supported. The fact that we have also learned something about the causes of A is a side benefit.

13It might be argued that this is not true in a court of law, where the consequences for the rock-dropper may differ depending on the court’s perception of the dropper’s intentions. But then the conditions are not the same. From a legal perspective, a rock dropped by accident versus with an intent to kill are two different events.

References


Borgatti and Halgin: On Network Theory
Organization Science, Articles in Advance, pp. 1–14, © 2011 INFORMS


Stephen P. Borgatti holds the Paul Chellgren Chair of Management at the LINKS Center for Social Network Analysis, Gatton College of Business and Economics, University of Kentucky. He received his Ph.D. in 1989 from the University of California, Irvine. His research interests include social network theory and methodology, knowledge management, and social cognition.

Daniel S. Halgin is an assistant professor of management at the LINKS Center for Social Network Analysis. He received his Ph.D. in management from Boston College. His research interests include social network theory, identity dynamics, and career trajectories.