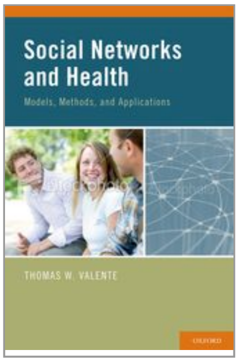


Social Networks and Health: Models, Methods, and Applications
Thomas W. Valente

Print publication date: 2010
Print ISBN-13: 9780195301014
Published to Oxford Scholarship Online: May 2010
DOI: 10.1093/acprof:oso/9780195301014.001.0001

Network-Level Measures

Thomas W. Valente

DOI:10.1093/acprof:oso/9780195301014.003.0008

Abstract and Keywords

Network level measures are measures calculated on the whole network and provide indicators of network structure. Eight network indicators are reviewed: size, density, reciprocity/mutuality, triadic census/transitivity, average path length, clustering, centralization, and core-periphery. The chapter then reviews two-mode data and closes with a discussion of network level influences on behavior. For example, how does density affect innovation diffusion?

Keywords: network level, density, network cohesion, network indicators

Network-level measures are measures calculated on the whole network. These measures provide indicators of network structure. Eight network indicators are reviewed: size, density, reciprocity/mutuality, triadic census/transitivity, average path length (APL), clustering, centralization, and core-periphery. The chapter then reviews two-mode data and closes with a discussion of network-level measures and behavior.

One of the more intuitive and interesting ways to study social networks is at the network level. The previous chapter discussed how a network could be reduced to a set of positions and the relations between positions were analyzed. Network-level measures examine the network globally, from a bird's eye view (if you will). Different networks within the same community can be compared. For example, how do friendship and advice-seeking compare in an organization? This chapter reviews various network-level measures and shows how network properties affect behaviors within those networks. Some might argue that because network measures ask about relations within the network, network-level analysis is the most appropriate level of analysis. The network-level measures will be presented, to some extent, in their order of complexity.

(p.129) Size

Network size is an important structural property, and it affects many other network indicators. Network size may be dictated by the study setting or research requirements. For example, studying adolescents in schools means that network size will be the classroom, grade, or school size. In other settings, the researcher may specifically set the network size of interest, perhaps by studying organizations with 100 to 250 employees. Some researchers estimate that a practical size for a group is about 150 people and some evidence exists that 150 is an optimal size for a human group (Dunbar, 1993).

Size may also be a function of individual network size or of how many people a person knows or has a meaningful relationship with. There is a practical limit on the number of people one can know or can maintain relationships with. Some researchers estimate the average size of acquaintance networks in the United States is approximately 280 (Killworth et al., 2006). A team of researchers from the University of Florida has conducted several studies to estimate network size (Killworth, 1990). Although people may know as many as several thousand other people, the set of contacts they may be able to name on any given topic is likely much smaller.

Size also indicates something about the networks of study. If one studies middle schools in various U.S. states and there are over 1,000 students in the middle schools of some states yet less than 50 students in the schools in others, it indicates something about the schools. Larger schools may have a larger range of student characteristics and a more complex administrative structure than smaller schools. Also, in larger schools, the range of possible people a student can develop relationships with is greater than in a small school. Size may not be the most interesting network indicator and may only reflect the boundary of the network, but it is the primary network indicator. As mentioned in Chapter 6, networksare often composed of separate components, in which case it is important to report the size of the components. The next network measure is density, the number of links in the network.

Density

Density is the number of connections in the network reported as a fraction of the total links possible. Density is calculated as:

8–1

$$D = l N (N - 1)$$

(p.130) where *l* is the number of links in the network and *n* is network size. Density is a key attribute of a network and should always be included as a covariate in analyses testing the effects of other structural properties. Equation 8-1 is applicable for asymmetric networks (in which the links are directed), but the numerator must be multiplied by 2 for undirected (symmetric) networks.

Density is also often calculated on subgroups within the network. For example, it might be hypothesized that network density differs by some attribute. Friendship networks among middle school students are often gender based so that boys are friends with boys and girls are friends with girls. Density can be calculated on ties among boys and between boys and girls. Density is a fundamental parameter for networks and subgroups within the networks. Density can also be used as a threshold value to determine interposition links as in blockmodeling (see Chapter 7).

There is an inverse relationship between size and density: as size increases, density decreases. Size is inversely related to density for at least two reasons. First, as mentioned earlier, there are practical limits to the number of other people a person knows or can establish relationships with. So as the boundary definition for the network increases, the limit on each individual's network size indicates that density will decrease. Second, in small groups, organizations, and communities, it is easier for people to know everyone in the network. For example, in a small organization of 10 employees, everyone will know everyone else and be pretty familiar with their day-to-day activities. Conversely, in a large organization of 250 or more employees, many employees will not know other employees. Consequently, all analysis of networks at the network level should include density as a covariate.

Although density is calculated as the proportion of links in the network, in nomination studies this formula may be modified to account for the limited number of network nominations solicited. In a nomination study in which the researcher asks respondents to name up to seven of their closest friends, the researcher may elect to report effective density, which is the number of links divided by size multiplied by seven. Thus, effective density is calculated as:

8–2

$$D E = l N (\lambda)$$

where *l* is the number of links, *N* is network size, and → is the maximum number of nominations requested.

Mutuality/Reciprocity

Reciprocated or mutual ties are links in which the direction goes both ways: If A chooses B, then B chooses A. Reciprocity is often referred to as *mutuality* because reciprocated ties are mutual, and this term may be preferred **(p.131)** because reciprocity can be a bit ambiguous. Reciprocity is ambiguous because there is both direct and indirect reciprocity (indirect reciprocity is when A chooses B but B does not choose A but chooses C who chooses A). Mutual, directly reciprocated, ties are symmetric while those that are not mutual are asymmetric.

Some relations (networks) are inherently mutual or symmetric, such as when asking, “With whom did you have lunch?” If a person had lunch with someone, it is expected that other person had lunch with the focal person. Conversely, other relations are inherently asymmetric such as, “Whom did you go to for advice?” It is quite likely a person goes to someone possessing greater expertise or authority for advice and that person is unlikely to reciprocate that nomination. Thus, measuring and understanding the degree of reciprocity in a network are important.

Reciprocity is measured (Borgatti et al., 2006):

8–3

$$R = (A_{ij} = 1) \text{ and } (A_{ji} = 1) \; (A_{ij} = 1) \text{ or } (A_{ji} = 1)$$

where A_{ij} indicates a link from i to j. A high degree of reciprocity indicates that people choose one another. This also means that people are choosing one another and not others and this can create more clustering within the network, thus increasing the distances in the network. Reciprocity can be an individual-level measure and a network-level one. At the individual level, reciprocity is a count or the proportion of reciprocated ties. At the network level, it is reported as the proportion of reciprocated links in the network (Equation 8-3).

Reciprocated ties may also suggest stronger ties. For example, in friendship networks, best friend ties are more likely to be reciprocated, whereas those among less close friends are less likely to be reciprocated. Tie strength can be measured as the order in which one provides names in response to a survey question, assuming that people provide the names of their closest friends first and less close friends later. Reciprocity can be included in this calculation by defining *closeness* as reciprocated friends.

For some behaviors, reciprocity may be associated with a greater likelihood of engaging in the behavior together or for the pair to influence each other. For example, several studies have shown that people are more likely to engage in risky behavior with closer ties than with distant ones (Valente & Vlahov, 2001). Risky drug and sexual behaviors occur more frequently with close and strong contacts because there is a higher degree of trust and intimacy with these close contacts and it would be impolite to use protective measures with such close friends. For example, two people engaged in a **(p.132)** romantic relationship often do not use condoms as protection because they want to communicate their trust and commitment to one another. In contrast, condom use may be more prevalent among casual sexual partners because there is no need to communicate trust and intimacy. In short, one might expect more behavioral influence between reciprocated relationships.

Triads/Transitivity

Reciprocity compared the links between two nodes, but what about the links between three nodes? The links between three nodes are referred to as *triads,* and in a directed network there are 16 possible combinations of links connecting three nodes (Holland & Leinhart, 1979). One measure of network structure is to describe the prevalence of these 16 types for the network (Figure 8–1). Of particular interest is transitivity in the network's triads. Transitivity in a network exists when there is the following combination of links among three nodes: if A → B, and B → C, then A → C (number 030T in Figure 8–1). This triad is considered transitive because A and B both have the same relationship to C.

Transitivity forms the basis of much sociological thinking about how people function in groups. Balance theory argued that people preferred a

(p.133) balanced environment with the people around them (Heider, 1958). If A and B are friends (from A's perspective) and A likes C, then A would want B to also like C. In balance theory, C could be a person or an attitude or object, such as a political opinion, a new product on the market, a behavior, and so on. People struggle to keep their world in balance, and Festinger (1954) introduced the idea of *cognitive dissonance,* which is the discomfort one feels when their environment is out of balance. Festinger (1954) argued that people will try to reduce their cognitive dissonance by trying to bring balance in their life—by trying to reduce intransitive triads.

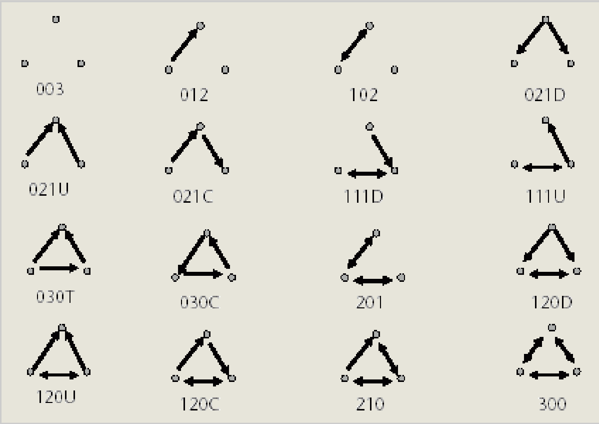


Figure 8–1. There are 16 possible configurations for the links between three nodes. These configurations are referred to as the MAN distribution for the number of mutual (M), asymmetric (A), and null (N) links among the nodes. Code 120 refers to 1 mutual, 2 asymmetric, and zero null links in the triad.

The tendency toward transitivity is the basis for the strength of weak ties theory proposed by Granovetter (1973). As Granovetter (1973) stated, the tendency for transitivity means that triads like number 201 in Figure 8–1 were uncommon (referred to as the forbidden triad [Granovetter, 1973]). So intransitive triads are rare, and because they are rare, there are few weak ties in the network. Yet the scarcity of these weak ties makes them strong in terms of their information capacity.

Researchers conduct transitivity analysis by reporting the proportion of transitive triads in the network. Networks with high levels of transitivity are thought to be cohesive and, thus, in a broad sense may be thought of as effective. Researchers conduct triad analysis by calculating the proportion of triads in the network that can be classified in the 16 categories Holland and Linehardt (1979) defined. These categories are often referred to as the MAN categories since there was a three-digit code defined for each category corresponding to the number of mutual (M), asymmetric (A), and null (N) links in each triad. So a code of 111 indicates one mutual tie, one asymmetric tie, and one null tie as the links among the three nodes, and a code of 210 indicates two mutual ties, one asymmetric, and zero null ones as the links among the three nodes.

So any network can be described in terms of the proportion of triads in the network of each of the 16 types. This is called a triad census. Understanding triads and their distribution in the network is important because triads have provided the basis for many social network effects (Faust, 2008). The strength of weak ties, for example, is predicated on the notion that there are few 201 triads (two mutual links and one null) and therefore weak ties are rare. Burt (1992) used triad census to stress the importance of structural holes, and Fernandez and Gould (1994) used triads to develop measures of brokerage. Faust (2008) has shown that using limited choice sociometric surveys (fixed number of nominations) result in networks that have a limited distribution of the triad census.

The reader might expect the next network structural property consists of analysis of four nodes, because reciprocity involved two nodes and triads/transitivity involved three. Analysis of the properties of four nodes, however, **(p.134)** is considered group analysis, and group calculations were discussed in Chapter 6. One network property in connection with groups is the number of components. It is important to determine the number of components in a network. The next network-level property concerns the number of steps it takes to transverse the network.

Diameter/Average Path Length

Two fundamental properties of a network are its diameter, the length of the longest path in the network, and the average distance between nodes. Networks with the same number of nodes, and even the same density (number of links), can have different diameters because the diameter is the number of steps in the longest path in the network. The network in Figure 8–2 shows who went to whom for advice in one community coalition designed to promote community-based substance abuse prevention activities (Jasuja, 2005; Valente et al., 2007). The length of the longest path in this network

(p.135) is five. For example, to go from person 3 to person 10 requires five steps: 3 → 26 → 17 → 2 → 5 → 10. There are numerous five-step paths in the network, and because this is the maximum in the network, it is the diameter. Network diameter then is the maximum distance between nodes in the network.

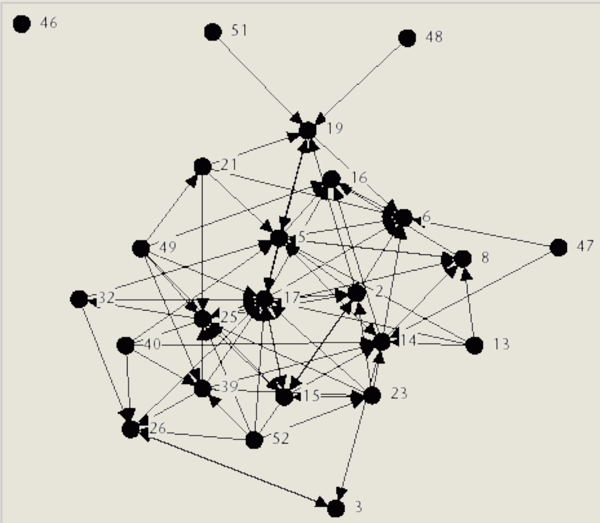


Figure 8–2. Network from one community coalition indicating to whom they turned for advice about prevention.

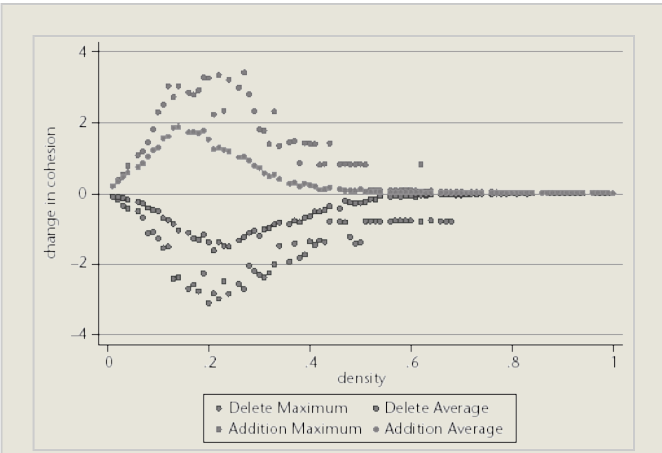
Average path length in a network is also an important structural property. The average path length (APL) is the average of the distances between all the nodes in a network. (APL is sometimes referred to as the characteristic path length.) A small APL indicates a cohesive network, while a large one indicates greater overall distances between nodes. For the network in Figure 8–3, the average distance between reachable people was 2.28. So, on average everyone in this network is 2.28 steps away from everyone else in the network provided they could reach one another. The APL is calculated by calculating the distances between all pairs of nodes in the network and then calculating the average (being sure to omit the diagonal).

Diameter and APL provide an overall indication of the structure of the network. Low diameter or low APL indicate a cohesive network with little clustering. The researcher need not know anything about specific group or positional distinctions but instead relies only on this one numeric indicator to draw conclusions about network topography. On the other hand, a network with large diameter and small APL may indicate that there are branches or spurs to the network that are mostly inaccessible to others in the network. One way to test the network structural tendencies would be to generate a sample, say 1,000, networks with the same size and density as the empirical (focal) network and calculate the diameter and APLs for these simulated networks (see Chapter 9). The researcher can then compare the empirical network indicators (diameter, APL) with the randomly generated distribution of networks with a test (a *t*-test) to determine if the network indicator is larger or smaller than would be expected by chance.

Density and Cohesion

All things being equal, as size increases, so does the APL, although not as much as one might think. On the other hand, the relationship between APL and density is nonlinear. As density increases, APL decreases, because high-density levels provide many paths along which to connect nodes. Research has shown, however, that the relationship between density and cohesion (APL) is not uniformly negative.

Figure 8–3 graphs changes in cohesion calculated for link deletions and additions for a large sample of randomly generated networks of size 10 for **(p.136)**



levels of density from 1 to 100 percent. At each density level, 50 different networks were created (because there are many different network structures at each density level). Notice that changes in cohesion (maximum and APL for the network) peak between 15% and 25% densities, indicating that networks in this density range can have their cohesion changed the most by the deletion of links. At 50% density and above, link changes have little effect on cohesion, because there are many redundant paths between nodes.

Figure 8–3. Maximum and average changes in cohesion scores as links are subtracted and added to the network by density. There is a sweet spot in social networks when density is between 15% and 25%, in which link changes have a profound affect on network cohesion. When density is very low and when it exceeds 50%, individual link changes have little effect on network cohesion because there is either no network (density is too low) or too much network (density is so high that individual links have little meaning).

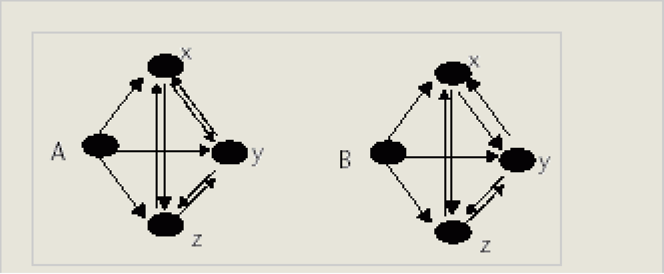
These simulated data indicate an interesting characteristic about networks. Density values above 50% indicate networks that have many redundant paths between nodes. Deleting links, or even nodes, in such networks will have little effect on overall network properties. Networks with 50% or greater density do not contain much structural information and in many cases are not likely to be interesting substantively. In such cases, it will probably be desirable to “prune” the network so that the researcher can find the hidden structure within the tangle of links. The pruning can be conducted several ways, perhaps first by considering stronger ties if the links are weighted. Next, the researcher can consider only reciprocated ties.

(p.137) Clustering

Network clustering is a measure of the degree of “clumpiness” in a network. A network with high clustering indicates that nodes are connected in dense pockets of interconnectivity, whereas one with low clustering has few pockets of interconnectivity. Clustering provides an index of the degree of structure in the network. So while the group and position analyses provided in Chapters 6 and 7 detail ways to find groups and positions, clustering provides a single measure of how much the groups and positions define or characterize the network.

Clustering is calculated by computing the average of each node's personal network density (Watts & Strogatz, 1998). Personal network density is the degree to which a person's ties are connected to one another. Figure 8–4 illustrates the calculation of personal network density—the person A on the left has three contacts (depicted as the thick solid lines, connecting A with x, y, and z). And x and z are connected to each other (as indicated by the heavier solid line connecting them). The five dashed lines indicate potential links among A's alters, but they are absent. So the personal network density for A is 1/6, one tie of six potential ones. In contrast, person B has three links connecting the people in his personal network, and so has a personal network density of 3/6, or 50%. The average personal network density for all the nodes in a network is the clustering coefficient.

A network with high average personal network densities indicates that people connected to a third person tend to be connected to each other. Notice this is similar to the triadic transitivity concept discussed in the previous section. Thus, transitivity indicates clustering. When personal network densities are low, transitivity is lacking and people are not more likely to know one another when connected to the same third party. Therefore, low



(p.138) personal network density indicates a lack of clustering. Another way to calculate clustering is to calculate the ratio of closed triads to all possible triads in the network (Luce & Perry, 1949; Wasserman & Faust, 1994).

Figure 8–4. Calculating personal network density. The solid lines represent links, and the dashed ones potential links. Personal network density is 1/6 for Bob, whereas it is 3/6 for Paul.

Clustering can accelerate behavior and disease spread within groups but will inhibit spread between groups. The average personal density in the network indicates the degree of clustering, and high rates of clustering indicate a network with pockets of interconnectivity separated by bridges that link these cohesive subgroups. The cohesive groups can be protective from a public health perspective as long as disease or risk behavior does not penetrate them. Once a disease penetrates these cohesive subgroups, however, it is likely to spread quickly, infecting everyone in a short period of time.

In clustered networks, then, how the clusters are linked to one another becomes critical. Two different types of network structures exist that link clustered subgroups: (1) a bridge structure in which the clustered subgroups are linked by bridges, and (2) a centralized structure in which the subgroups are linked by central nodes who occupy powerful positions in the network. The best indicator to distinguish these two types of structures is centralization.

Centralization

Centralization is the degree a network's ties are focused on one person or a set of people. In centralized networks, one or a few people hold positions of power and control in the network, whereas decentralized networks have diffuse power and control structures. Figure 8–5 shows two pairs of networks with the same densities, yet different centralization scores. The networks on the left are decentralized (e.g., centralization = 9.1%), while the ones on the right are centralized (e.g., 50.9%). It is important to stress that density and centralization are not correlated and are independent structural measures of the network. A dense network can be decentralized or centralized, and a sparse network can be decentralized or centralized. The centralized networks on the right (one dense and one sparse) have a node at the center of the network that receives a disproportionate amount of ties.

Centralization was discussed in Chapter 5 in reviewing centrality measures. Recall that there were numerous individual centrality measures and each one has a network-level counterpart. For example, centrality degree measures the number of links a person sends and receives. Degree centralization degree (NB: centraliza*tion* for the network level) indicates whether one or some nodes send or receive a disproportionate number of links and can be calculated on in-degree and out-degree. Centralization is calculated **(p.139)** by determining the maximum individual centrality score in the network and subtracting it from all other individual scores in the network. These differences are summed and that total is divided by the maximum sum of differences theoretically possible in a network of that size (which happens to be given conveniently by a formula). For example, the formula for centralization degree (CD) is (Freeman, 1979):

8–4

$$CD = \frac{\sum (Max(C_{Di}) - C_{Di})}{n^2 - 3n + 2}$$

where $Max(C_{Di})$ is the maximum centrality score in the network, C_{Di} indicates the individual centrality scores, and *n* is network size. A similar **(p.140)** formula exists for centralization based on centrality closeness, betweenness, and others. Another way to calculate centralization is by simply calculating the standard deviation of the centrality scores for the network. A large standard deviation indicates a lot of variation in the individual centrality scores, whereas a small one indicates little variation and hence a decentralized structure.

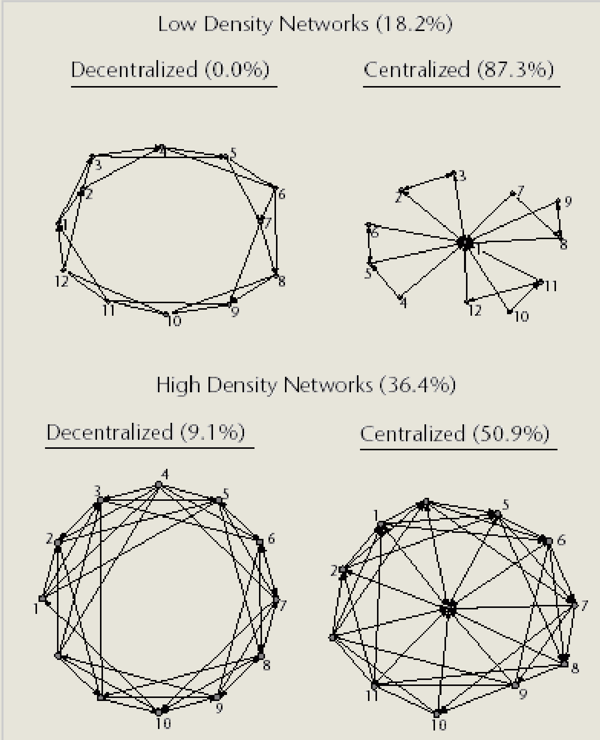


Figure 8–5. The top two networks have equal network density, 18.2%, yet the one on the left is decentralized (centralization = 0%) and the on the right is centralized (centralization = 87.3%). Similarly, the bottom two networks have equal densities, 36.4%, yet dramatically different centralizations (9.1% and 50.9%, respectively).

With nomination data, in-degree centralization will usually be greater than out-degree centralization. Typically, nomination data limit the nominations the respondent can provide. This limits the number of out-degree nominations, and so the variance in out-degree is smaller, making out-degree centralization smaller. In contrast, the nominations received distribution can be quite varied because one person or a few people will receive many nominations and so the out-degree distribution is greater. This, of course, depends on the data and the extent of actual centralization in the network.

Centralized networks have the potential for fast diffusion because people at the center of the network can act like broadcasters and reach many people quickly. Conversely, centralized networks may have slower diffusion because the central nodes are gatekeepers and have a disproportionate amount of influence in the network. If the central nodes delay adoption or if they do not support the new idea, then diffusion will be slowed in centralized networks. Thus, centralization has the net effect of placing more power and control on the central people, the opinion leaders. Opinion leaders are more critical to behavior spread in centralized networks rather than in decentralized ones.

The net effect, then, of centralization may be on the shape of the diffusion curve rather than on how long it takes for diffusion to reach saturation. Centralized networks will have steeper diffusion and the time until diffusion takes off rapidly will depend on the attitudes and behaviors of the central people. If the central people are resistant, then the diffusion will remain at a low level until leaders embrace the new idea and then it will increase rapidly. Conversely, if the central people embrace the idea early, then diffusion will accelerate rapidly. A more nuanced version of centralization is the idea of a core-periphery structure.

Core-Periphery

Many empirical networks exhibit a core-periphery structure. Core-periphery structures are networks in which there is a group of nodes who are densely connected to one another (the core) and a separate group of nodes loosely connected to this core and loosely (or not at all) connected to each other (the periphery). Core-periphery networks may have somewhat low centralization scores because the people in the core all have similar centrality scores. Yet, **(p.141)** there is still considerable structure in the network but it is not a centralized structure, or a clustered structure, rather a core-periphery one.

The degree to which a network has a core-periphery structure is determined by fitting a core-periphery model to the data (Borgatti & Everett, 1999) by permuting the data matrix repeatedly such that nodes are alternatively in the core or periphery. The various empirically based core-periphery structures are correlated with an idealized core-periphery structure (connections among core nodes and with peripheral nodes and no connections between peripheral nodes). The best model is the one in which node assignments in the data have the maximum correlation with the idealized core-periphery structure. This correlation is the fit index indicating how well the data conform to a core-periphery structure (Borgatti & Everett, 1999).

Figure 8–6 displays a collaboration network that has a modest core-periphery structure typical of coalition networks (Valente et al., 2008b). The fit index for this core-periphery structure was 0.362, which is the correlation coefficient between the observed data matrix and a hypothetical ideal core-periphery matrix (Borgatti & Everett, 1999). The fit index can be interpreted as a correlation coefficient, indicating that these data only modestly conform to a core-periphery structure. Core organizations in Figure 8–6 (n = 18) are

(p.142) depicted as circles and peripheral ones (n = 13) as squares. In this study, there were many different types of organizations in the core of the network including government, health plans, school, philanthropic, academic, policy, and community-based organization/advocacy. Thus, the coalition was not dominated by one or a few organizations or organization types; rather collaboration was shared by many different types of organizations. It was also the case that being in the core was a function of organizational members’ meeting attendance, involvement in coalition activities, and a higher percentage of the organization's goals being related to children (Valente et al., 2008b).

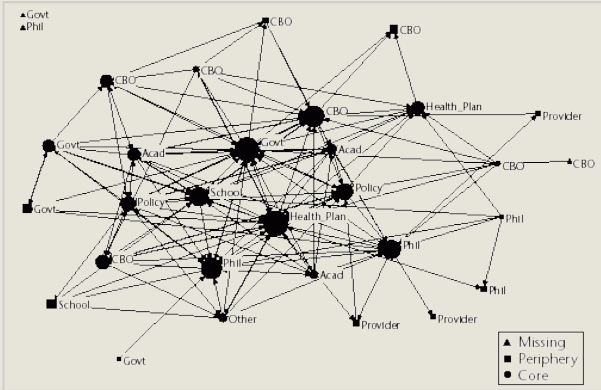


Figure 8–6. The network of collaboration among 31 organizations that participated in a coalition to expand children's health insurance. Circles are core organizations, and squares are peripheral ones. Nodes are sized based on in-degree in the collaboration network and positioned using spring embedding.

These are the primary indicators used to describe network structure: number of components, size, density, reciprocity, triad census (especially transitivity), APL (distance), clustering, centralization, and core-peripheriness. Table 8–1 reports the averages, standard deviations, and ranges for these network indicators for 24 communities from the STEP study. These are baseline network data derived from asking coalition members to name the other members they turned to for advice on prevention issues (Valente et al., 2007). The average size was 22.04 members, ranging from as few as 4 to as many as 41. There was an average of 12.7 components in the networks, and all of these consisted of one large component with each isolate being considered a separate component. The average density was 13%, with 20% of the ties being reciprocated. Some 37% of the triads were transitive, with one network having all triads transitive.

The APL in the largest connected component was 2.2, ranging from 1 to 3.1. This indicates that most coalition members could reach everyone else in the coalition (if they could reach them at all) in three or fewer steps. The maximum path length was 5.1, ranging from 1 to 8. The clustering coefficient

Table 8–1. Selected Metrics for 24 Coalition Advice Net-works

	Mean	SD	Range (Min, Max)
Size	22.04	8.58	4, 41
Density	0.13	0.06	0.06, 0.33
Proportion of reciprocated ties	0.20	0.08	0.05, 0.40
Proportion of transitive triads	0.37	0.16	0.19, 1.00
Average geodesic distance*	2.16	0.45	1.00, 3.14
Maximum geodesic distance	5.08	1.64	1, 8
Number of component	12.67	5.56	3, 24
Clustering coefficient	0.31	0.13	.15,.67
Norm. In-degree centralization (%)	34.63	11.27	16.63, 54.50
Norm. Out-degree centralization (%)	18.06	7.89	8.41, 44.44
Number of isolates	2.92	8.58	0, 8

(*) Computed among reachable nodes.

(p.143) average was 0.31, indicating moderately clustered networks. The networks were more centralized on in-degree (34.6%) and then on out-degree (18.1%), which is common in nomination studies because most respondents provide a similar number of nominations given the restriction on the number allowed and yet people/organizations can receive as many nominations as there are others in the network. There was an average of nearly three isolates per coalition.

As mentioned earlier, network indicators vary with one another and with outcomes. For example, size was negatively associated with density: the larger the network, the less dense it is (r = –0.68). Size was also negatively associated with the percentage of ties reciprocated (r = –0.55), the percentage transitive (r = –0.63), and clustering (r = –0.59). Size was positively associated with average path length (r = 0.80). Thus, in this study, the larger the network, the more likely that people do not reciprocate nominations and form tightly nested triads in which everyone knows one another. Density was also associated with other metrics. For example, density increased reciprocity (r = 0.48) and transitivity (r = 0.80) and decreased the APL (r = –0.69).

One thing to note is that the APL will vary significantly depending on how disconnected nodes are treated. For the 24 networks in this study, the mean APL was 2.16 (SD = 0.45) when measured only among connected nodes in the largest component. The APL would increase dramatically if N – 1 was used as the distance between disconnected nodes to 18.9 (SD = 8.22) and 5.60 (SD = 1.57) when D + 1 was used where D equals the maximum distance between connected nodes. The lay public has become quite familiar with the small world concept associated with the term “six degrees of separation” (see Chapter 1). The measure of “six degrees” is the APL, and to report APL in a way that is consistent with “six degrees” concept one would consider only connected nodes, although in some instances it may be preferred to calculate network properties with the inclusion of disconnected nodes.

It is also possible to determine if network-level characteristics are associated with network-level outcomes such as how successful the coalition has been at achieving its objectives. These network-level outcomes can be derived from external indicators (Did the coalition achieve extramural funding?) or from aggregates of individual responses (How effective do coalition members report the coalition as being”). In the STEP study, coalition density was negatively associated with adoption of evidence-based practices (Valente et al., 2007), whereas centralization was positively associated adoption of evidence-based practices (Fujimoto et al., in press).

One would expect networks with shorter overall path lengths to be more cohesive and conducive to the adoption of new behaviors. Because shorter APL indicates a shorter distance between everyone in the network, **(p.144)** information has to travel less far to reach everyone and so diffusion should be facilitated. The tendency toward reciprocity and transitivity, however, will lengthen overall path lengths because it creates clustering in the network.

Two-Mode Data

In Chapter 3, five different types of network data collection were provided and the notion of two-mode data was introduced. Two-mode data are data derived from information on events, organizations, or situations in which people participate. For example, one might have data on the departments to which the employees in an organization belong. Two-mode data are arrayed in a table referred to as an *affiliation matrix*, in which people are the rows and the events or organizations are the columns. The events (columns) should be binary vectors in which a 0 means the person does not belong or did not attend the event and a 1 means the person attended or belonged. Table 8–2 provides an example from the class network shown in Figure 7–1. The students in this class were from different departments within the university. (The affiliation matrix in Table 8–2 is also sometimes referred to as an *incidence matrix*.)

The data show that every student belonged to one department, and one faculty member belonged to two departments (was jointly appointed). The two-mode data can be used to construct a person-by-person network based on shared departments by multiplying the Table in 8–1 by its transpose. The transpose of a matrix is the matrix turned on its side. So this affiliation matrix has 27 rows and 4 columns (ignoring the row and column labels) and its transpose has 4 rows by 27 columns. These two matrices can be multiplied because they are conformable (the number of columns in the first matrix equals the number of rows in the second one) and the resulting matrix has 27 rows and columns (is a 27-by-27 square matrix, which are the outside dimensions of the two matrices).

The product of these two matrices indicates which people are in the same department and the network can be graphed (Figure 8–7). The network shows connections between all of the people in the same department. There is one link connecting the two departments, the faculty with appointments in the two departments. The network has been constructed from information only about who belonged to which department. A matrix (network) of department-by-department affiliations can also be constructed from this same data. The cells of the department-by-department matrix indicate the number of students who share departments. The department-by-department matrix is created by multiplying the transposed matrix by the original one (reversing the order of the multiplication) so the first matrix is 4 × 27 and the second is **(p.145)**

Table 8-2. Departmental Affilia-tions of the Students (27 × 4 Matrix)

	Dept. A	Dept. B	Dept. C	Other
A_S1	1	0	0	0
PD_F1	0	1	0	0
A_S2	1	0	0	0
A_S3	1	0	0	0
Oth_F2	0	0	0	1
B_S4	0	1	0	0
A_S5	1	0	0	0
A_S6	1	0	0	0
A_S7	1	0	0	0
A_S8	1	0	0	0
A_S9	1	0	0	0
B_F2	0	1	0	0
B_S5	0	1	0	0
A_S7	1	0	0	0
A_S8	1	0	0	0
B_S4	0	1	0	0
C_S1	0	0	1	0

	Dept. A	Dept. B	Dept. C	Other
B_S5	0	1	0	0
C_S2	0	0	1	0
A_S9	1	0	0	0
A_S10	1	0	0	0
B_S6	0	1	0	0
A_F3	1	1	0	0
C_S3	0	0	1	0
A_S11	1	0	0	0
B_S7	0	1	0	0
A_S12	1	0	0	0

27 × 4. Thus, the affiliation data provide two networks that can be graphed and correlated with attributes or other networks.

For example, it might be hypothesized that department affiliation was associated with the network of who-knew-whom when the class started. The correlation between the department affiliation matrix and the choices of who knew whom is calculated using a technique known as the Quadratic Assignment Procedure (QAP) available in statistical programs (Krackhardt, 1987, 1988; Borgatti et al., 2006). QAP provides a means to correlate to two matrices that accounts for the dependencies in the matrix.

(p.146)

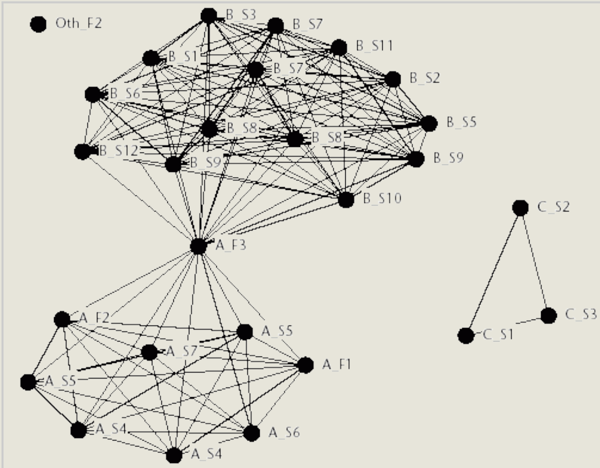


Figure 8–7. Graph of two-mode data, in which links indicates people who belonged to the same department derived by converting an affiliation matrix (department membership) into one mode adjacency matrix.

In this case, the “know” network shown in Figure 7–1 was regressed on the network derived from departmental affiliations using the two-mode technique (Figure 8–7). The standardized regression coefficient (?) was 0.23 ($p < .001$), indicating that department affiliation was weakly associated with who knew whom at the start of this class. The correlation was not as high as expected, although it is statistically significant.

Individual Network-Level Interactions

One theoretically interesting avenue of research is the interaction between individual-level and network-level measures. These interactions can be of two types. First, interactions between individual measures such as centrality can be studied in the context of a network metric such as centralization. For example, are central members of an organization more likely to be successful in a centralized organization than a decentralized one. In this case, the **(p. 147)** overall network structure is hypothesized to affect the interpretation of the effects of the individual metric.

The second type would be an interaction between a network-level metric and network influence. For example, it might be hypothesized that social influence is stronger in dense networks than in sparse ones. Dense networks have more connections and so more opportunities for information to flow, and they may be indicative of a norm in which communication occurs. Conversely, dense networks may be “noisy” and so social influence for each individual competes with others in the focal person's personal network. Therefore, the other hypothesis is that social influence is stronger in sparse networks. One way to test this proposition would be to compare network exposure based on the number of ties who engage in the behavior compared to the same measure divided by the number of ties (comparing network exposure as a count versus a proportion). Another way would be to use data from multiple networks with identical measures so that social influence can be compared across sparse and dense networks.

One might also think that social influence varies by network position. For example, are people on the periphery of the network more likely to be influenced by peers than those at the center? Because people in the center have more connections or occupy a privileged position in the network, they can be more selective about how others influence them. People on the periphery, in contrast, may only learn about behaviors from one person and so may be more susceptible to that person's influence. Given the plethora of network level and individual level measures, the number of hypotheses is enormous.

Summary

This chapter described the network level of analysis in network analysis. Network-level indicators are those measures which describe the entire network. Size is the most basic network measure, and network density is a fundamental property. Mutuality, or the extent to which ties are reciprocated, is also a network level property and indicates whether there is a tendency for ties in the network to be reciprocated. Network researchers also examine the way three nodes are connected. Triads, three nodes, can be connected in as many as 16 different ways and researchers have documented these ways in what is known as a triad census. Every network can be described by the frequency the triads in the network occur in the 16 different possible permutations. Of particular interest are transitive triads, those in which A → B and B → C implies A → C.

A useful network metric is the APL of the network. APL is often used as a measure of network cohesion in which networks with low APL are **(p.148)** considered cohesive. Networks with an APL greater than expected by chance are considered fragmented. The chapter showed how APL and density correlate. Another network indicator is analysis of whether the network has a core-periphery structure. The chapter included a discussion of two-mode data and how it is used to create networks of people-by-people and group-by-group. The chapter also provided an example of two-mode data and how it can be correlated with other networks to answer substantive questions. The chapter closed with a discussion on network-level influences on behavior.