Structure in personal networks[*]

Christopher McCarty
PO Box 117145
Bureau of Economic and Business Research
University of Florida
Gainesville, Florida

Voice: 352.392.0171.332
Fax: 352.392.4739

E-mail: ufchris@ufl.edu

Abstract

Most personal (egocentric) network studies describe networks using measures that are not structural, opting instead for attribute-based analyses that summarize the relationships of the respondent to network members. Those researchers that have used structural measures have done so on networks of less than 10 members who represent the network core. Although much has been learned by focusing on attribute-based analyses of personal network data, the application of structural analyses that are traditionally used on whole (sociocentric) network data may prove fruitful.   The utility of this approach becomes apparent when the sample of network members elicited is relatively large.

Forty six respondents free-listed 60 network members and evaluated tie strength between all 1,770 unique pairs of members. Graph-based measures of cohesion and subgroups revealed variability in the personal network structure.   Non-hierarchical clustering generated subgroups which were subsequently verified by respondents as meaningful. Further analysis of the correlation between subgroup types and overlap between subgroups demonstrates how the analysis of each network can be summarized across subjects. Four case studies are presented to illustrate the richness of the data and the value of contrasting individual matrix results to the norm as defined by all 45 subjects.

Keywords: Personal network, Egocentric network, Social network structure

Introduction

Personal (egocentric) network studies have as their focus the social relations of individuals. Typically, the researcher elicits the names of family, friends and acquaintances from respondents who have no relationship to each other. These data are analyzed by summarizing the characteristics of the network members for each respondent and correlating aggregate characteristics with the individual characteristics of the respondents themselves. Data like these enable the researcher to study topics such as the characteristics of social support networks (Wellman and Wortley, 1990), variations in social support following a disaster (Beggs, Haines and Hurlbert, 1996), network influences on gun-carrying behavior among Black adolescents (Myers, et al., 1997) the relationship between IV drug use and the transmission of HIV (Neaigus, et al., 1994), or the effect of personal networks on voting behavior (Nieuwbeerta and Flap, 2000).

In contrast, those who study whole (sociocentric) networks are interested in the pattern of relations among respondents who form a socially defined group. The group might be members of a club, a classroom of children, or the executive board of a Fortune 500 company. Those who study whole networks measure the strength of tie between all members of the group, and thus have a proximity matrix that represents their pattern of relations. They apply a set of matrix-based analytic techniques, some unique to social networks (such as centrality and density) and others that are commonly used in multivariate statistics (such as cluster analysis and multidimensional scaling). Researchers who use the whole network approach examine such topics as state formation in Renaissance Italy (Padgett and Ansell, 1993), the structure of Japanese corporate networks (Gerlach, 1992), or corporate innovations (Raider, 1998).

Among social network researchers, it is often assumed that the application of analyses typical of whole networks do not apply to personal networks. As it turns out, with some exceptions, there are no mathematical or statistical reasons that prohibit the application of these matrix techniques to personal network data. There are, however, logistic and conceptual constraints.

First, generating matrix data for individual respondents is time-consuming and potentially expensive. Data about a respondent's personal network are generated by the respondents themselves, thus they themselves must assess the tie strength between every unique pair of network members. For a network size of 50 people that is 1,225 pairs; for a network of 100 people that is 4,950 pairs.

Although this is a tedious task, individual assessments are often easy to make. People tend to classify network members into groups, and frequently members of one group do not know members of another. For example, it is not uncommon for there to be no network ties between family and co-workers. Assessments of network ties between these network members takes little time, particularly if the evaluation is for the presence or the absence of a tie, and not for tie strength. This study will show that respondents can make such evaluations in a reasonable time and, in most cases, with little fatigue.

The second obstacle is a related issue, that is the fundamental assumption that a respondent can report accurately about the members of their network. Many network researchers question the accuracy of a respondent's report about their ties with others, let alone their report about ties between network members. Critics would claim that at best we are working with "perceptions" of relations, rather than actual relations.

Much has been written about the value of measuring social perception versus observable relations (see Mayhew, 1981), and the debate will not be resolved here. My own view is that at least some portion of behavior is driven by perceptions of the environment, physical or social, and that a respondent's report of ties reflect the social environment. Further, I believe that respondents *can* report accurately about relations between network members, particularly if they are reporting only the presence or absence of a relation. Again, this study will show that the structures derived from personal network data are meaningful to respondents, lending support to the claim that their assessments of ties are accurate.

Finally, in a whole network study, the goal is to derive structure, each network member contributing equally to that structure. Some network members will be more central than others, but they are central by virtue of the contribution of all network members. The outsider in the group contributes to the definition of another network member's more central position.

In contrast, the personal network approach is designed to determine the influence of each network member on the respondent. This approach relies on the fact that each network member does *not* contribute equally to that respondent's personality and behavior. Yet the matrix approach assumes that all network members contribute equally to the structure. Ego's co-worker contributes as much to the structural properties of the network as ego's mother. A structural analysis will not account for the unique contribution of a single network member. We can, however, see how attributes of the network members are associated with structure.

Typical personal network data

Campbell and Lee (1991) and McCarty, et al. (1997) summarize the types of data that are collected in personal network studies. Nearly all of the measures compiled from the data are summaries of attributes of network members that are then compared to the same or other attributes of respondents. Socio-demographic variables such as age, sex, and race are often collected for network members, and these variables are then summarized as averages or percentages. For example, one might compare the average age of network members and see how this varies by the age of the respondents.

Other typical questions asked of respondents concern the type or content of relation the respondent has with each network member. The relationship may be classified by the types of things the respondent expects the network member to do (Bernard, et al., 1990), or a classification of their relationship within a set scheme (Fischer, 1982; Bernard et. al., 1988; McCarty et al., 1997). In all cases, the categories of relations are created by the researcher.

Nearly all personal network researchers collect some measure of tie strength; that is, how close the respondent is to each network member they list. Questions may be very explicit, such as "how often do you see X" or "how long have you known X". Questions may also be more abstract, such as "how well do you know X". (McCarty, 1996, Marsden and Campbell, 1984).

All of these data are attributes of the network member, and so far attributes have not been used to study structure. It may be possible to analyze network structure using attributes of network members by using differences between attributes as a proxy for similarity or dissimilarity. A tie between two network members could be approximated

by their similarity on a given attribute, such as race, sex, age or a combination of these. It is safe to say that structure derived from these definitions would look very different from structure derived from a respondent's assessments of ties between network members.

For these reasons personal network researchers usually do not ask respondents to evaluate the existence of a tie between network members. Those who have asked such questions limit them to a small set of core network members who, as core members, often have a high degree of interaction. Because of this close interaction, the analysis of structure among core network members demonstrates little variance between respondents (Marsden, 1990).

As I said, the study reported here is not the first attempt to apply structural measures to personal network data. Network density has been measured in several personal networkstudies (see Wellman, 1979; Fischer, 1982; Burt, 1984; Campbell and Lee, 1991; Volker, 2001). The difference between past studies and the one reported here has to do entirely with the larger number of network members and the addition of the tie evaluations between each member. With a relatively small number of network members who tend to be close friends and family, there is very little structural variation to explain. When the canvas of potential relationships is broadened to include many types of network members, acquaintances as well as friends, a different looking structure is revealed. And for many behaviors we would expect the effects of structure to come from indirect ties as well as ties that are closer and more direct (Walker, Wasserman and Wellman, 1993; Gottlieb, 1981; Wellman, 1981).

The most thorough analysis of structure in personal networks is that of Burt and his work on structural holes (1992). Burt uses his own measure of redundancy in

personal networks as a proxy for variation in "brokerage opportunities". He then relates

this variation to respondent-level variables such as personality (1998), or to power and

mobility within a corporate network. Burt has amassed a large database of ego networks

of managers ranging from 6 to 20 alters, that he uses as a benchmark for management

consulting. While some researchers have written on the concept of structural holes

(Krackhardt, 1999) or issues with its measurement (Borgatti, 1997), very few have

actually collected personal network data and applied Burt's concepts, and none, including

Burt, have done so outside of a business context.

Spreen (1992, 1999) and Spreen and Zwaagstra (1994) have derived methods for

estimating personal network structures using sample data. Their research is part of an

effort to derive the structure of large whole networks, particularly that of hidden or

elusive populations such as heroin users. This tradition, pioneered by Frank (1978) relies

on snowball sampling on known members of these populations to estimate the structural

characteristics of the whole network. The focus of these research efforts have been on

estimating the structure of the large whole networks where the boundaries are unknown,

and not on the description or analysis of the structure of the personal network themselves.

The Data

These data originate from a study of perceived subgroups in personal networks

that stemmed from questions about the validity of relation categories used in personal

network research. Typically researchers create categories then ask respondents which

category best describes their tie to the network member (e.g. family, co-worker, religious

affiliation). The purpose of this research was to have the respondent free list the names

of network members, and evaluate the ties between all members. The goal was to create

clusters based on perceived interactions of network members that might reveal biases in the preconceived relation categories that are often used in personal network research.[1]

Forty six respondents were recruited through friends and newspaper ads. Respondents were asked to "list 60 people you know -- that is, people who are alive, who you can recognize by sight or by name, who recognize you by sight or by name, and who you can contact if you had to." Respondents were given no examples of the type of person they could include, but were allowed to use nicknames if they could not remember exact names. The 46 respondents produced 2,760 names. After naming 60 alters, respondents indicated how well they knew each one (on a scale of 1 to 5 where 1 was "do not know very well" and a 5 was "know very well").

They then described, in their own words, how they knew each alter ("she's my sister," "she's the daughter of my professor," "he's my former boss," "he's the drummer in my friend's band," "she's my ex-boyfriend's ex-girlfriend"). Respondents were also shown a list of 23 ways in which people know each other and were asked to check off up to three of these ways in which they knew each of their 60 members.

Finally, respondents were presented with each of the 1,770 unique pairs of their 60 members and were asked, for each pair: "Do these two people know each other, and if so, then on a scale of 1-5, how well do they know each other?" This part of the data collection exercise took 45 minutes, on average. To test the reliability of this procedure,

---

[1] Lorrain and White (1971), extending on Nadel's theories made similar arguments for complete networks, suggesting that by examining the pattern of individual relations, network roles would emerge. The theory and extensive work on structural equivalence follows from that paper (Burt 1982, Krackhardt 1999).

respondents coded every 59th pair a second time. For these 30 pairs, 93% were coded identically the second time; 97% were coded as knowing one another at a scale level ±1 from the original coding.

Structural analysis using graph-based measures

Unlike statistical measures, such as multidimensional scaling and cluster analysis, graph based measures rely on the presence or absence of ties between alters and ignore the strength of tie. In order to use these measures the data must be recoded. Thus, if a respondent evaluated the tie between two alters as greater than or equal to 1 then it was coded as a 1, otherwise it was coded as a 0. For convenience, I have provided definitions of the six graph-based measures calculated on the 46 proximity matrices. The summary measures are provided in Table 1.

Density – The percent of ties that exist in a network out of all possible ties. A density of 1.0 implies that every alter is connected to every other alter. A density of 0 implies that no alter knows any other alter.

Degree Centrality –A measure of network activity. An alter is highly degree central to the extent that he or she is directly connected to many other alters. A graph (network) is highly degree central to the extent that there is wide variability of point degree centrality among alters. A star network, where one alter is the intermediary for all other alters, would be 100% centralized. A network where all alters have the same number of ties within the network would have graph degree centrality of 0. Degree centrality is a measure of direct ties, and thus should be used for concepts that require direct ties.

Closeness Centrality – A measure of independence from the control of others. This is a similar concept to degree centrality, except that it focuses on the path rather than direct ties alone, so an alter is still be considered to be connected (reachable) through intermediaries. A single alter is highly close central if they are connected by short paths to many other alters. Like degree centrality, 100% centralization implies a star network and 0 implies that all alters have the same number of ties. Closeness centrality in an egocentric network would be a useful measure when the concept being studied does not require a direct tie.

Betweenness Centrality – A measure of information control. A single alter is highly between central to the extent that they lie on many geodesics (shortest paths) between

alters. In this sense they act as a bridge between alters, and thus potentially control information. Again, 100% centralization of the entire network implies that one alter is a bridge to all others, as with the star network, and betweenness centralization of 0 implies that no alter is any more a bridge than any other alter, such as in a circle graph.

Cliques – A clique in an egocentric network is a set of alters who are all directly tied to each other. There can be overlap between cliques, that is an alter can be a member of more than one clique. Since cliques are maximal complete graphs, you cannot make a clique from a subset of clique members. Therefore the number of cliques that exist in a graph is a measure of the number of subgroups that exist.

Components – A component in an egocentric network is a set of alters who are connected to one another directly or indirectly. Unlike a clique, the members of a component do not have to be connected to everyone else in the subgroup. If there is a path to an alter, they will be a member of the component. A network with many components implies a compartmentalized network.

All of the measures listed in Table 1 are sensitive to the number of alters in the network. Since respondents were required to list exactly 60 alters, the measures in Table 1 are all comparable. This underscores the value of constraining egocentric network elicitation to the same number of alters when structural comparisons are to be made.

The first thing that is apparent from Table 1 is that structure within personal networks varies. For example, closeness centrality ranges from 0.44 to 84.7, where the possible range is 0 to 100. The coefficient of variation (the standard deviation divided by the mean) indicates that of the six measures, closeness centrality varies the most among the 46 respondents.

Table 2 shows the relationship between these graph measures, nearly half of which are significantly correlated. By examining how these measures relate to each other it becomes more clear what they mean and how they might be applied in network analysis. For example, there is a strong negative association between components and density. This is because, in a highly dense network, people tend to be connected to each other by at least one path, thus limiting the number of independent components.

Similarly, a highly dense network tends toward many people being potentially connected to each other, which leads to low closeness centrality, resulting in a strong negative correlation. Alternatively, a highly dense network creates the opportunity for many subgroups where everybody knows each other (a clique), resulting in a positive correlation between network density and the number of cliques.

 The correlation between density and degree centrality is not significant. Given that degree centrality is focused on direct ties, as is network density, this may seem surprising.   Density, however, is simply a measure of ties within the entire network while centrality is a measure of cohesion within the network. This points to the value of centrality scores and the limitation of network density as a measure of cohesion.

The above analysis shows that graph centrality varies significantly across respondents' personal networks, and that the three types of centrality (degree, closeness and betweenness) focus on different aspects of the network. Table 3 is based on an analysis of the most point-central alter in each network for each type of centrality. The first row shows that 44% of the respondents had the same alter as the most point central alter for the three separate measures. In contrast, 11% had different alters for each centrality measure. One explanation for the different patters within these networks is the geographic separation of alters. Among those respondents where there is a different alter for each type of centrality, on average nearly half were from some place outside the state of Florida (where the data were collected), compared to less than a third for those where the same alter was the most central for all three measures.

Interviews from subsequent studies using a similar format suggest that personal network structure may reflect different network strategies. Some people tend to

compartmentalize their network alters, consciously attempting to keep them in separate groups, while others try to bring people together as much as possible. One can easily imagine consequences of these different strategies in terms of information flow, control and social support, and so exploring this variability, perhaps with personality inventories, might be a fruitful area for research.

Cluster analysis

Statistical analyses make use of the variability in tie strength, and thus the raw tie evaluations were used for the 46 matrices submitted to the cluster analysis. However, to ensure that the groups were formed on ties that truly did exist, the matrices were recoded to binary data where a tie of 4 or 5 was recoded as a 1 and all other ties were recoded as 0. These matrices were then cluster-analyzed to search for subgroups based on respondents' perception of the strength of ties rather than preconceived categories of relations.

The clusters were computed using non-hierarchical rather than hierarchical clustering. With hierarchical clustering members can belong to only one cluster. With non-hierarchical clustering, members can belong to more than one cluster (Arabie et al. 1981). As implemented in the Statistical Analysis System (SAS), the ADCLUS procedure requires the user to specify a maximum number of clusters. After some trial runs, it appeared that there was diminishing variance explained for most respondents after 14 clusters were extracted. Ultimately 14 was established as the cutoff for the number of clusters.

Interpreting clusters is a subjective exercise. Using the information respondents gave about each network member, I tried to determine why the members of the cluster

would know each other. Even though this process allowed for more variability in the types of clusters than most personal research, it became apparent that other clusters might be identified if I had asked respondents for other details about their network members. As is the case with any qualitative interviewing, the information is often a function of the amount and direction of probing.

After coding the 45 sets of clusters, there seemed to be 12 major categories, three of which were further divided into a total of 19 cluster types. Table 4 shows the cluster types and the percentage of respondents who had clusters corresponding to those types. With up to 14 clusters extracted by ADCLUS for each respondent, there could be up to 14 of the 19 cluster types represented in each case. This never occurred. The minimum number of cluster types was 3 and the maximum was 11. On average, respondents had 6.3 (sd 1.8) cluster types.

All respondents had clusters composed of family members. For some respondents, the family clusters were clearly composed of either paternal or maternal kind, and for about half the respondents (49%) there were clusters composed of the spouse's (or significant other's) family. Respondents who had paternal family clusters tended not to have maternal family clusters, and vice versa. Six respondents had maternal family clusters and seven respondents had paternal family clusters. Only two respondents had both maternal and paternal family clusters. It appears that among these respondents, one side of the family is favored (or recalled) over the other.

There was a clear set of family clusters based on the respondent's having labeled the cluster members as "close." In most cases this was congruent with the level of

knowing assigned by respondents to the cluster members. I labeled these clusters as "close family." "General family" included both close and distant family.

More than half of the respondents (56%) had clusters whose members were tied to a particular intermediary. Clusters via other people and clusters composed of couples are closely linked types. Most network vias were composed of a couple and their family, usually where the respondent knew one member of the couple strongly.

Twenty-one (46%) of the respondents had clusters that were composed of people with whom they had gone to school. For 40% of the respondents school clusters were made up of members who were known from the local university, while just 11% of respondents had clusters of members known from high school.

Finally, there were clusters composed of persons who shared a religious affiliation or a common hobby, or who were neighbors or members of a social group, or who shared a common issue. Examples of this last type were clusters of persons who were in the same therapy or encounter groups or who were members of political or group rights organizations. These clusters were not mutually exclusive. Members could be in more than one cluster type; that is, clusters overlapped.

To determine the validity of the clusters, and particularly to test my labeling of those clusters, a reliability check with ten of the original respondents was done. Each respondent was presented with a card that had either 1) a cluster produced by his or her own data, or 2) a card containing one of seven clusters generated by selecting random numbers between 1 and 60 to fill clusters of 2, 4, 6, 8, 10, 12 and 14 members. Respondents saw seven random clusters each (of sizes 2, 4, etc.) as well as the clusters generated from their data.   They were asked to label the group of people in some way.

All clusters formed by the clustering procedure were recognized by respondents as an intuitively clear subgroup. Furthermore, respondents described the contents of all clusters with labels that agreed with my labels. Only six random clusters out of 70 were identified by respondents as meaningful.

Correlation of cluster types

It is reasonable to assume that there are some cluster types that occur together and others that do not. Table 5 shows the results of a correlation analysis of the presence or absence of cluster types by respondent. The first row indicates that those who associate with a significant other's family, most likely a spouse, have more close family to form close family clusters. The negative correlation between close and general family suggests that there is a tendency for those with close family not to have general family, rather than having both close and general family. There is a negative correlation between close family and hobby groups.

College groups are negatively associated with neighbors, religious groups and current work. College students are less likely than non-students to have full time jobs, or to be heavily involved in church groups. Given their transitive living conditions, they are less likely to stay in the same neighborhood for several years, and thus less likely to form social groups made up of neighbors.

The negative association between high school groups and former job is clearly related to time. Those who are close enough to high school that they maintain those groups are less likely to have had the experience of having a former job.

A factor analysis of this correlation matrix yielded very similar results. Perhaps the most revealing result from the factor analysis was the first factor which loaded close

family against general or distant family. In personal network research there is a tendency to classify network members who are family together. These data show that, among these respondents, there is a distinction between family members that may be as important as distinctions between network members based on the function of the relation.

Overlap

There are few studies that examine overlap within personal networks. Some whole network studies have described the overlap between network members and events (Galaskiewicz, J. and P. Marsden , 1978). Milardo (1989) studied the overlap within personal networks generated by two different methods, but the analysis was not structural. The following analysis is intended to demonstrate how members of a personal network occupy more than one relational subgroup.

Over the 39 respondents who listed members who overlapped across cluster types, there were on average 6.4 (sd 4.7) overlapping members per respondent. Overlapping members exhibit some characteristics which distinguish them from non-overlapping members. They are significantly (.003) older with an average age of 39 versus 36, and have known the respondent for significantly (.001) longer than non-overlapping members (20 years compared to 12 for non-overlapping members). Overlapping members rank significantly (.001) higher on the knowing scale than non-overlapping members (4.5 compared to 3.4).

Table 6 makes clear the dominance of the family categories for overlapping members. In other words, members who tend to exist in more than one cluster type, also tend to exist in some type of family cluster. The dominance of family clusters in the overlap is certainly a function of the dominance of family members in the respondent's

list and clusters as a whole. While overlap between clusters is a common occurrence, it is not clear that it often occurs with subgroups that are not alike (such as work and religion clusters).

Recalculation of clusters

These findings are intriguing pointers to possible rules about how network subgroups are held cognitively by people in the US. A better test requires a much larger and more representative sample of respondents. The procedure for collecting these data, however, takes too much time and effort (by both respondents and interviewers) for wide application. If substantially similar results could be extracted from a much smaller sample, then the method might be useful in large field surveys, or even in telephone interviews.

With hindsight, some of the detail about network members was not needed for categorizing clusters. Those details could be eliminated in a field or telephone survey, cutting down on the time it takes to administer this procedure. The true time savings will come from reducing the total number of network members involved in the process.

I repeated the clustering procedure with data from five respondents using the first 50, 40, 30 and 20 members they had named on their free lists. All five respondents were able to generate some clusters at 30 members or more. Neither the number of clusters nor the number of cluster types consistently dropped with shorter lists. Indeed, in one case the number of clusters rose.

Of greater interest are two indices that weight the ability of the reduced member sets to replicate the number of clusters and cluster types using all 60 members -- presumably a more accurate representation of the total network. These indices are the

number of valid clusters divided by the number of network members and the number of cluster types divided by the number of network members. In four out of five cases, both indices peaked at the 30 member level, suggesting that 30 is an optimal number of members for discovering clusters. Reducing the procedure to a list of 30 members would cut the length by more than half since the generation of the matrix by respondents would consist of judgments about 435 pairs rather than 1,770 pairs.

However, the tendency toward order-effects associated with free-listing suggests that a subset of members should be randomly selected from a larger free listed set (Brewer, 2000). A shorter version of this instrument would have respondents free-list 60 members, randomly select 30 from the list of 60 and conduct the data collection on this reduced set.

Case studies

The preceding analysis has shown that respondents recognize groupings of their members which are not necessarily intuitive and which could elicit a different list of members were they to be used in a network generator. The categories listed above are subjective generalizations of some very specific groupings listed by respondents. Yet the following multidimensional scaling plots of four respondents' adjacency matrices, which they themselves helped label, lend credibility to the subjective labeling and demonstrate the potential for this type of analysis .

Steve

Steve's case history highlights the effects that location plays in the formation and maintenance of subgroups. Originally from Australia, Steve pursued an MBA in Toronto, Canada and, at the time of data collection, was working on a Ph.D. at the

University of Florida. He maintained social contact with his graduate cohort and with faculty members and belonged to several hobby groups (see Figure 1).

The members in Steve's list are more similar in education than are the members in many other lists, and Steve named more female members than males did on average (52% versus 43%). One of the questions I asked all respondents about each of their listed members was whether a change in location would terminate the relation after a year. On average, respondents said that 29% of their members would drop out under these circumstances. Steve said that only 3 of his 60 members (5%) would drop out of his list after not seeing them for a year.

Steve's centrality measures are all three far below the average, implying less structural cohesion than other respondents. At 13.03, his betweenness centrality score is among the lowest of the 46 respondents, indicating the absence of bridges between the subgroups in his network. Steve is one of the few respondents who has a different alter as the most point central for the three centrality scores. His most between central alter isa friend who accompanied him on his trip from Toronto, where he was an MBA student, to Gainesville, Florida where he became a graduate student. This provided one of the few links between Steve's many subgroups.

Steve generated more sensible clusters than most respondents; 12 clusters compared to an average of 10. He also generated more cluster types than the average; 7 compared to 6.3. These higher levels are explained by his history of living in many places combined with an active social life.

Steve's cluster types reflect his involvement in a Ph.D. program. Of the 7 cluster types used, the university accounted for five clusters. On the other hand he had no close

family clusters at all. In contrast to other respondents, Steve does not have a strong core of interacting family members. Steve said that his mother is a focal point for he, his brother and sister, but there are few occasions where several family members get together. The strong presence of hobby and social groups comes second only to the high number of university related groups.

This reflection of dispersed and unrelated groups is perhaps most noticed in the area of overlap. Steve is below the average with only 2 of 60 respondents crossing cluster types, compared to 6.4 (sd 4.7) over all respondents. This fits his mercurial social strategy, moving from place to place, group to group, retaining cores of group members.

Betty

In contrast to Steve who is white, male, and highly mobile, Betty is a black female university secretary with a high school education and strong ties to her home area. Unlike Steve who has traveled extensively, Betty has rarely traveled outside of Florida and Georgia. She has a strong sense of belonging to the area and close family ties.

Betty keeps her private life separate from her work life. When not at work she prefers to socialize with her family and only rarely participates in social functions with co-workers. She visits her family in Georgia often.

Betty names more females than males (62% to 38% respectively) in her network, many more than the 56% female respondents named on average. In terms of maintenance of the relationship, Betty has no examples of situational relations, a significant difference (.001) from the group average.

Her mean knowing level was significantly higher (.001) than the average over all 2,820 members, at 4.1 compared to 3.5. Oddly, she says she would lose contact with a

much higher than average percentage of her members than the group would if she moved away. Betty's relationships are location-bound.

Betty had the lowest closeness and betweenness centrality scores of any of the respondents. This reflects the fact that most of her network consists of two massive groups (work and family) that are highly connected, but with virtually no connections between them.

Betty had a total of 12 clusters and seven cluster types. Her current work accounted for half of the clusters, while close family accounted for two. She did not include her husband on the list despite his important role in her life. Again, this points to the potential problems associated with free-listed recall.

Unlike Steve, Betty has a high degree of overlap between clusters, and across cluster type. These are almost entirely accounted for by the overlapping of work and social group clusters. Figure 2 demonstrates the importance of work relations in Betty's list of 60 members. The right side of the plot is dominated by her current work. Far on the left is her family which shares no links to her work, and her in-laws at the bottom who are equally separated from her biological family.

Tony and Mary

Tony and Mary met in high school and dated for many years, and were married when they graduated from college. They moved to the Orlando area after graduating, and worked as a two income family for several years. Both Tony and Mary had family in a town very close to Orlando and had several groups, frequently couples, with whom they socialized. Eventually, Mary was accepted to the dental program at the University of

Florida. They moved back to Gainesville where Tony accepted a position at the university.

Dental school is rigorous, particularly in the earlier years. Mary's 12 to 16 hour days left little time, or opportunity, for establishing relationships outside the dental school routine. This left less time for socializing with Tony.

Like Steve, Tony is a gregarious person who enjoys socializing with others. Whenever Mary was available, he spent his time with her, and because of her obligations to socialize with the dental organizations, he attended many of their functions. Yet these dental functions were not an everyday occurrence. Eventually Tony developed several social activities independent of Mary.

Tony had four components, twice the average among the 46 respondents. Again, like Steve, Tony tends to try to maintain relationships with his alters even after he moves. This results in subgroups with little or no connection, such as his family in Kentucky and his former co-workers in Orlando. It also results in some unintuitive centrality results, such as the fact that his most degree and close central alter is his uncle's wife. This demonstrates the problems with graph-based measures in that they cannot incorporate the strength of tie and thus might inflate a particular alter's position.

Mary used exactly half male and half female members, which is a much higher percentage of males than female respondents used on average. This is no doubt due to the high proportion of males in her dental school cohort.

Mary's degree and closeness centrality scores are much higher than average, again reflecting her immersion in the dental program where most students knew each other. Because there are few bridges to make, her betweenness centrality score is lower

than average at 19.54. Her most degree central and close central alter is a faculty member within the dental program.

With six clusters, Mary is over one standard deviation below the group average. This is again due to the dominance of the dental school cluster which explained 69% of the total variance in the cluster analysis. There is little room for the expression of other clusters as the dental cluster accounts for over half of Mary's members and there is a high degree of inter-knowing within that cluster. Her one family cluster is a mixture of close and distant family, with no tight-knit core.

Tony is very close to the group average for many network characteristics. He has no cases where someone else maintains the relation, but significantly (.001) more incidence of situational relationships, more than double the average. He has strong family ties and similarly strong representation from work and former work. Indeed, it seems that wherever Tony works, he knows people he maintains as a subgroup. These are probably social groups he has been involved with for many years.

Mary and Tony's MDS charts are interesting (see Figures 3 and 4). On both charts they are each other's closest member. In Mary's chart the dental cluster with 40% of her members, the Orlando grouping and the grouping of business associates are easily distinguished. Even though she is highly embedded in the dental cluster, her relations to the other clusters involve Tony. In fact, Tony is integrated into the dental cluster by virtue of his attendance at the many dental school socials where spouse participation is encouraged. Similarly, Tony's clusters of family, work and former work are equidistant from he and Mary in the center. She is a member of the clusters as well, largely through him.

Discussion

This study demonstrates the potential of analyzing the structural aspects of personal networks. It is not unique in concept, as others have examined network density among core network members. Although free-listing of network members presents some problems with recall and the exclusion of weak ties (Brewer, 2000), expanding the list of network members to 60 from the typical 5 to 10 yields a more representative sample of the total personal network. Others have suggested that by including more distant ties, structural phenomena may emerge that explain respondents' behaviors in several areas, particularly social support (Milardo; 1989; Walker, Wasserman and Wellman, 1993).

It is clear from these data that graph based measures of structure do indeed vary across respondents. One interpretation of this variability is that it reflects strategies of respondents in terms of compartmentalizing their networks and in time investment. With a more heterogeneous sample and the inclusion of additional measures, analysis of these structural measures may be useful for predicting respondent attributes, such as scores on a personality inventory or or a scale of depression.

Many of the categories that emerge from this analysis are unsurprising. I expect, for example, to find that people naturally categorize their network members into family, neighbors, people they know from work, people they went to (or go to) school with, and people with whom they share a hobby. I would, in fact, be skeptical if I did not find these results.

On the other hand, 42% of the respondents distinguish between people they know from their current job and people they know from some former workplace; 25% of members were named because they belong to the respondents' networks via some

intermediary network member. Neither of these findings is intuitive, but it is precisely because of the intuitive findings that I have confidence in the surprising ones.

One of the reviewers of this article brought up an important point, suggesting that the structural analysis of the personal network doesn't tell us much about the respondent than we could get by asking them general qualitative questions about the people they know. Would such an interview generate the finding of a negative correlation between close family and hobby clusters? Would I find that clusters of network members tied to a particular intermediary are as prevalent as they appear to be?

It is unlikely that any person could make these assessments of their networks qualitatively. Trends in other structural characteristics that I did not report on in this paper, such as centrality and density, would be virtually impossible to detect without a systematic analysis such as this. Certainly more analysis of these data can, and will, be done that complement what one can learn simply from talking to someone about the people they know.

The case studies allude to a practical application of this method. One can imagine a therapist using the MDS plots, clusters and the relational data as a vehicle for analyzing the "health" of a client's support network. This assumes there are certain properties of social support networks which can be agreed upon as healthy or unhealthy. Already some studies have suggested clients are more satisfied with the existence of confidantes (Conner, Powers and Bultena 1979), certain density levels (Hirsch 1979, 1980), or levels of network size (Polister 1980).

Finally, this study has shown how personal network matrices generated from several respondents can be analyzed across respondents to reveal personal network

characteristics that have been ignored until now. The fact that cluster types were independently verified by a sample of respondents demonstrates the validity of the inter-network member tie evaluations. I have shown that the network characteristics derived from these matrices can be summarized across respondents to show trends in the composition of the networks.

What separates this paper from other studies of personal networks is nothing more than the extra time and effort of asking respondents about the relationship between all potential pairs in a large personal network. While this is tedious for respondents, computer software can make this task easier by taking advantage of patterns that emerge as the respondent evaluates pairs. The development of such software dedicated to the structural analysis of personal networks, and the aggregation across respondents' networks, would be a powerful tool.

References

Arabie, P., J.D. Carroll, W. DeSarbo and J. Wind 1981. "Overlapping Clustering: A New Method for Product Positioning." *Journal of Marketing Research* 18:310-317.

Beggs, J.J., V.A. Haines and J.S. Hurlbert 1996. "Situational Contingencies Surrounding the Receipt of Informal Support." *Social Forces* 75:201-222

Bernard, H.R., P.D. Killworth, M.J. Evans, C. McCarty and G.A. Shelley 1988 "Studying Social Relations Cross-Culturally." *Ethnology* 27:155-179.

Bernard, H.R., P.D. Killworth, C.McCarty and G.A. Shelley 1990. "Comparing Four Different Methods for Measuring Personal Social Networks." *Social Networks* 12:179-215.

Borgatti, Stephen P. 1997 "Structural holes: Unpacking Burt's redundancy measure" *Connections* 20:35-38.

Brewer, D.D. 2000. "Forgetting in the recall-based elicitation of personal and social networks." *Social Networks* 22:29-43.

Burt, R. 1982. Toward a structural theory of action : network models of social structure, perception, and action. New York : Academic Press.

Burt, R. 1984. "Network Items and the General Social Survey." *Social Networks* 6:293-339.

Burt, Ronald S. 1992 Structural Holes. Harvard University Press: Cambridge

Burt, Ronald S., Joseph E.. Jannotta, James T. Mahoney 1998 "Personality correlates of structural holes" *Social Networks* 20:63-87.

Campbell, K.E. and B.A. Lee 1991. "Name Generators in Surveys of Personal Networks." *Social Networks* 13:203-221.

Conner, K.A., E.A. Powers and G.L. Bultena 1979. "'Social Interaction and Life Satisfaction: An Empirical Assessment of Late-life Patterns." *Journal of Gerontology* 34:116-21.

Fischer, C. S. 1982. *To Dwell Among Friends*. Chicago: University of Chicago Press.

Frank, O. 1978. "Sampling and estimation in large social networks" *Social Networks* 1: 91-101.

Galaskiewicz, J. and P. Marsden 1978. "Interorganizational Resource Networks: Formal Patterns of Overlap. *Social Science Research* 46:89-107.

Gerlach, Michael L. 1992. "The Japanese Corporate Network: A Blockmodel Analysis." *Administrative Science Quarterly* 37:105-34.

Gottlieb, B. H. 1981. "Preventive Interventions Involving Social Networks and Social Support." in *Social Networks and Social Support*, edited by B.H. Gottlieb Newbury Park, CA: Sage.

Hirsch, B.J. 1979. "Psychological Dimensions of Social Networks: A Multidimensional Analysis." *American Journal of Community Psychology* 7:263-78.

Hirsch, B.J. 1980. "Natural Support Systems and Coping with Major Life Change." *American Journal of Community Psychology* 8:159-72.

Krackhardt, David 1999 "The ties that torture: Simmelian tie analysis in organizations" *Research in the Sociology of Organizations* 16: 183-210.

Marsden, P.V. and K.E. Campbell 1984. "Measuring Tie Strength." *Social Forces* 63:483-501.

Marsden, P.V. 1990 "Network Data and Measurement." *Annual Review of Sociology* 16:435-63.

Mayhew, Bruce. 1981. "Structuralism vs. Individualism: Part I, Shadowboxing in the Dark" *Social Forces* 59: 335-375

McCarty, C. 1996 "The Meaning of Knowing as a Network Tie' *Connections* 18:20-31.

McCarty, C., H.R. Bernard, P.D. Killworth, E.C., Johnsen and G.A. Shelley 1997 "Eliciting Representative Samples of Personal networks" *Social Networks* 19:303-323.

Milardo, Robert M. 1989. "Theoretical and Methodological Issues in the Identification of the Social Networks of Spouses." *Journal of Marriage and the Family* 51:165-174.

Myers, G.P., G.A. McGrady, C. Marrow, C.W. Mueller 1997 "Weapon carrying among Black Adolescents: A Social Network Perspective." *American Journal of Public Health* 87:1038-1040.

Neaigus, Alan, Samuel R. Friedman, Richard Curtis, Don C. Des Jarlais, R. Terry Furst, Benny Jose, Patrice Mota, Bruce Stepherson, Meryl Sufian, Thomas Ward, Jerome W. Wright 1994. "The Relevance of Drug Injectors' Social and Risk Networks for Understanding and Preventing HIV Infection." *Social Science and Medicine* 38:67-78.

Nieuwbeerta, P. and Henk Flap 2000. "Crosscutting social circles and political choice: Effects of personal network composition on voting behavior in The Netherlands." *Social Networks* 22:313-335.

Padgett, John F., Christohper K. Ansell 1993. "Robust Action and the Rise of the Medici, 1400-1434." *American Journal of Sociology* 98:1259-1319.

Polister, P.E. 1980. "Network Analysis and the Logic of Social Support." in *Evaluation and Action in the Environment,* edited by R.H. Price and P.E. Polister. New York:Academic Press.

Raider, Holly J. 1998. "Market Structure and Innovation." *Social Science Research* 27:1-20.

Spreen, M. 1992. "Rare populations, hidden populations, and link-tracing designs: What and why?" *Bulletin de Methodologie Sociologique*. 36: 34-58.

Spreen, M. and R. Zwaagstra 1994. "Personal network sampling, outdegree analysis and multilevel analysis: Introducing the network concept in studies of hidden populations." *International Sociology*. 9: 475-491.

Spreen, M. 1999 Sampling personal network structures: Statistical inference in ego-graphs. Unpublished dissertation, University of Groningen.

Walker, Michael E., Stanley Wasserman, Barry Wellman 1993. "Statistical Models for Social Support Networks." *Sociological Methods and Research* 22:71-98.

Wellman, B. 1979. "The Community Question: The Intimate Networks of East Yorkers." *American Journal of Sociology* 84:1201-1231 .

Wellman, B. 1981 "Applying Network Analysis to the Study of Social Support." in *Social Networks and Social Support*. Edited by B.H. Gottlieb Newbury Park, CA: Sage.

Wellman, B. and S. Wortley 1990. "Different Strokes from Different Folks: Community Ties and Social Support." *American Journal of Sociology* 96:558-588.

Table 1. Mean values for six graph-based measures.

Measure	What it measures	Mean	Minimum	Maximum	Standard Deviation	Coefficient of variation
Density	Ties	0.24	.11	.56	0.09	36.32
Degree	Cohesion	46.05	19.18	80.07	17.27	37.51
Closeness	Cohesion	33.60	0.44	84.7	25.11	74.73
Betweenness	Cohesion	29.18	1.65	57.8	14.26	48.88
Cliques	Subgroups	74.30	21	204	39.41	53.04
Components	Subgroups	1.93	1	6	1.25	64.83

Table 2. Correlation of graph-based measures -- Value(Prob > r)

	Density	Degree	Closeness	Betweenness	Cliques	Components
Density		.07(.632)	.41(.005)	-.41(.005)	.49(.001)	-.45(.002)
Degree			.56(.001)	-.30(.042)	.12(.425)	-.29(.053)
Closeness				.27(.065)	.22(.139)	-.72(.001)
Betweenness					.30(.042)	.07(.623)
Cliques						-.22(.131)
Components						

Table 3. Pattern of same point central alter across three types of centrality.

Close - Degree	Close - Between	Degree - Between	Percent
No	No	No	11
No	No	Yes	15
No	Yes	No	15
Yes	No	No	15
Yes	Yes	Yes	44

Table 4. Percent of respondents who had a given cluster type.

Cluster type	Frequency	Percent
Family	45	100
Maternal	6	13
Paternal	7	16
Close	32	71
General	28	62
Significant other's family	22	49
Including friends	25	56
Network via other person	25	56
Couples	23	51
Childhood, growing up	4	9
School together	21	47
High school	5	11
University	18	40
Work together	33	73
Current work	23	51
Former work	19	42
Housemates	3	7
Religious affiliation	5	11
Hobby group	6	13
Issue oriented group	8	18
Neighbors	13	29
Social group	13	29

Table 5. Correlation of the presence of cluster types.

Cluster type 1	Cluster type 2	Pearson's R	Probability > R
Close Family	Significant Other's Family	.43	.01
Former Work	Family, Including Friends	-.32	.03
Hobby	Issue Oriented	.33	.03
Hobby	Close Family	-.33	.03
General Family	Network Via	-.33	.03
College	Neighbor	-.32	.03
Hobby	General Family	.31	.04
Housemates	Former Work	.31	.04
General Family	Religion	-.31	.04
High School	Former Work	-.30	.04
Issue Oriented	Network Via	.30	.05
Close Family	General Family	-.29	.05
College	Religion	-.29	.05
College	Current Work	-.29	.05
Housemates	Network Via	-.30	.05

Table 6. Frequency of overlapping cluster types.

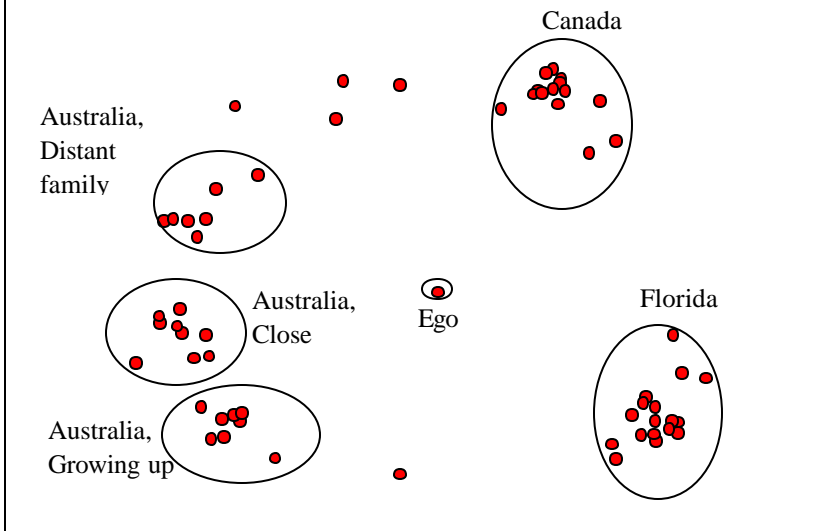| Category 1	Category 2	Frequency
Close Family	General Family	16
Close Family	Significant other's family	15
Close Family	Family including friends	15
Close Family	Network via other person	15
General Family	Family including friends	12
Close Family	Couples	11
Family including friends	Network via other person	11
General Family	Significant other's family	9
Close Family	Former work	8
General Family	Network via other person	8
General Family	Social group	8
Significant other's family	Network via other person	8

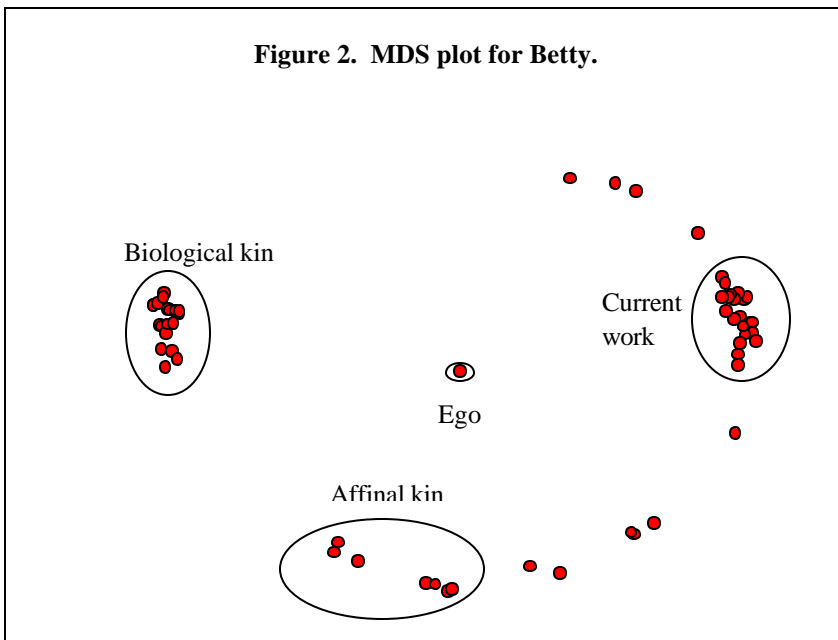Figure 1. MDS plot for Steve

Figure 2. MDS plot for Betty.

Figure 3. MDS plot for Mary.

Family in Orlando
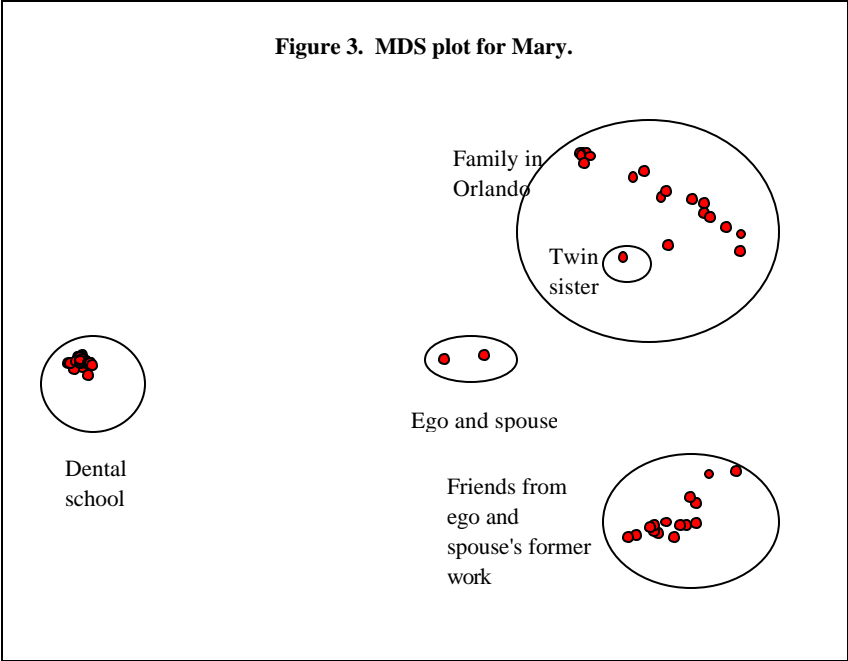
Twin sister

Ego and spouse

Dental school

Friends from ego and spouse's former work

Figure 4. MDS plot for Tony.

Current work

Biological family

Ego and spouse

In-laws

Former work #2

Former work #1