Published in final edited form as:
Nor Epidemiol. 2009 ; 19(1): 5–16.

Social Network Visualization in Epidemiology

Nicholas A. Christakis, M.D., Ph.D., M.P.H. and
Department of Health Care Policy, Harvard Medical School; Department of Medicine, Mt. Auburn Hospital, Harvard Medical School; and Department of Sociology, Harvard University

James H. Fowler, Ph.D.
Department of Political Science, University of California, San Diego

Abstract

Epidemiological investigations and interventions are increasingly focusing on social networks. Two aspects of social networks are relevant in this regard: the structure of networks and the function of networks. A better understanding of the processes that determine how networks form and how they operate with respect to the spread of behavior holds promise for improving public health. Visualizing social networks is a key to both research and interventions. Network images supplement statistical analyses and allow the identification of groups of people for targeting, the identification of central and peripheral individuals, and the clarification of the macro-structure of the network in a way that should affect public health interventions. People are inter-connected and so their health is inter-connected. Inter-personal health effects in social networks provide a new foundation for public health.

A person with more friends and social contacts generally has better health than a person with fewer friends,[i,ii,iii] and a person at the center of a network is more susceptible to both the benefits and risks of social connection (e.g., for infectious disease) than those at the periphery of a network.[iv,v,vi] People are thus affected by their location in a social network. In addition, and distinctly, they are influenced by behaviors and outcomes in people who are “nearby” them in the network (including their friends, friends of friends, and so on). It is not just how connected a person is, but also who a person is connected to, and what those people are doing, that has an effect. Indeed, social networks affect health through a variety of mechanisms, including: (1) provision of social support, (2) social influence (*e.g.*, norms, social control), (3) social engagement, (4) person-to-person contacts (*e.g.*, pathogen exposure), and (5) access to resources (*e.g.*, money, jobs, information).[vii] New work with social networks suggests that such interpersonal effects extend beyond just those individuals to whom a person is directly connected. Health-related phenomena, whether germs or information or behaviors, can diffuse widely within social networks.

The scientific objectives in social network analysis are generally 1) to understand the processes that determine the topology, or structure, of the network, and 2) to understand the extent and mechanisms behind any inter-personal effects within the network. Social network analysis also promises to provide targets for intervention, by identifying influential individuals, by identifying cliques of at-risk individuals, or by elucidating procedures for maximizing the impact of health interventions. Creating visual images of networks can serve important heuristic purposes in both research and policy, and visual images are powerful complements to quantitative analyses.

Correspondence: Nicholas A. Christakis, Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave., Boston, MA 02115, (617) 432-5890, christakis@hcp.med.harvard.edu.

Social Network Elements and Attributes

Social networks consist of two elements: individuals (nodes) and the social ties between them. Once all the nodes and ties are known, one can draw pictures of the network and discern every person's *location* within it, placing each individual in social space analogous to geographic space. Within a network, one can speak of the "distance" between two people (also known as the "geodesic distance" or "degree of separation"), which is the shortest path in the network from one person to another. For example, a person is one degree removed from her friend, two degrees removed from her friend's friend, three degrees removed from her friend's friend's friend, and so on. Social network ties are not restricted to friends, of course, and one may be connected to one's spouse's brother's friend, or one's co-worker's friend's sister, and so on. In discussing network effects, it is helpful to refer to "egos," or the individuals under study, and their "alters," or the people to whom they are connected (though the same person may be an ego and an alter from different perspectives).

Social ties may be described as "edges" (undirected relationships between nodes) or "arcs" (directed relationships from one node to another). Examples of undirected relationships include spouses and siblings. A directed relationship is one such as between two friends in which A identifies B as a friend but B does not reciprocate. Edges and arcs are often measured on a binary (presence/absence) scale, but may also be valued (e.g., how well two people know each other or how much they like each other).

Network data may be fruitfully represented in matrix or graphical form. A matrix indicates relationships between every person and every other person by coding numbers indicating the existence or nature of relationships in a square table. A graph shows these connections visually, as illustrated by Figures 1–4. Here, we focus on so-called sociocentric network studies. These are studies that attempt to discern all the inter-connections between members of a defined population. In contrast, an "egocentric" study collects information about people's contacts from each respondent, without the requirement that the alters also be in the study.

A social relationship is "transitive" if there is redundancy between direct and indirect relationships. If A knows B and A knows C, then their relationship is transitive if B and C also know one another. High transitivity is a fundamental property of human social networks. If we were merely making random connections with people drawn from the population, then there would be an extremely low probability that two of a person's friends are also friends with one another. Instead, we frequently make friends with our friends' friends, thus dramatically increasing the probability of transitive relationships.

Often, real social networks contain collections of sub-networks or "components." A component is a part of a network in which everyone is connected by at least one path to every other person in the same part. Logically, this means that for two different components, no one in the first component can be connected to anyone in the second component. If the largest component contains a majority of nodes in a network, it is known as a "giant component." Analyses of graphs also leads to the identification of subgroups of nodes that are more closely linked to one another in "communities" or "cliques" than they are to the rest of the network.

In its simplest form, network analysis focuses on connections between homogenous nodes, but most substantively interesting sets of network data incorporate information on the characteristics of nodes, such as their gender, income, or health behaviors. Other characteristics distinguishing nodes from one another may be based on examination of the network itself. For example, in a network of undirected ties, the *degree* of a node is its number of other nodes to which it is directly connected. For a directed graph consisting of

arcs, the number of arcs that point to a node is that node's *in-degree*; the number of arcs that emanate from a node is its *out-degree*. Nodes can obviously vary in their degree. Nodes having higher degree are usually interpreted as being more prominent and influential within the network.

Measures of *centrality* in networks capture the extent to which a node connects, or lies between, other nodes, and hence its tendency to be positioned near the center of his or her local network. Centrality is also taken as a marker of importance and prominence. The simplest measure of centrality is the aforementioned count of the number of friends (known as "degree" centrality). People with more friends will tend to be more central. But this measure does not account for differences in the centrality of one's friends. Individuals who are connected to many well-connected peers are more central than those who are connected to an identical number of poorly-connected peers. In other words, those who befriend popular people should be more central than those who befriend unpopular people. "Eigenvector centrality" captures this.[viii] This measure assumes that the centrality of a given subject is an increasing function of the sum of all the centralities of all the subjects to whom he or she is connected. Eigenvector centrality values are inherently relative: an individual connected to every other person in the network would have the maximum possible value, and a person not connected to anyone else would have a value of 0. Other measures of centrality capture the extent to which a particular node lies *between* other nodes, i.e., the extent to which that node lies along geodesic paths linking other nodes and hence must be traversed if something (*e.g.*, information, germs, money) is to move between other nodes in the network.[ix] Of course, with visual images of networks, it is quite straightforward to see who is in the middle of the network and who is on the periphery.

Finally, a so-called one-mode network includes a single type of node, such as patients. In a two-mode or "bipartite" network, there are two types of nodes (e.g., physicians and patients) and all relationships link nodes in one set to nodes in the other set (e.g., patients 1, 2 and 3 see doctor A, and patients 3, 4 and 5 see doctor B). A two-mode network can be reduced to a one-mode network on the basis of indirect relationships: physicians are connected to one another indirectly through their shared patients, and patients are connected to one another by virtue of having physicians in common.

Basic Network Topologies

Networks often have stereotypical structures which have distinct visual appearances. For example, they might be organized in forms such as regular lattices, small-world networks, scale-free networks, or random graphs. In regular lattices, nodes are highly clustered, having ties only with their nearest neighbors. At the other extreme, in a random graph, nodes are randomly linked together, and there is no local clustering; geodesic paths between any two nodes in random graphs are relatively short. Small-world networks display substantial local clustering, but include a small number of "shortcuts" to distant parts of the network; like a random graph, they have short path lengths between any two nodes.[x] These shortcuts reduce the time its takes for "communication" from one end of the network to the other, and this property can dramatically change disease transmission dynamics.[xi] In a scale-free network, ties between nodes do not occur purely at random but rather may result from a process of preferential attraction, in which nodes having more links are disproportionately apt to acquire new ones.[xii] A scale-free network displays a so-called power law in its degree distribution (i.e., the distribution of the number of ties each node has); it has far more nodes of high degree than would be found in a random graph. These high-degree nodes function as hubs and the short path length found in random networks is found in scale-free networks. Still other network topologies are also possible. For example, in ring networks, the nodes are largely arrayed along large loops). Naturally occurring human networks do not

necessarily conform to the foregoing types, in part because of the complexity of the underlying processes which create them.[xiii]

Drawing Networks

Once a full set of individuals and the ties among them are observed, there is only one network *per se*. However, this network can be analyzed or drawn in various ways. For example, when drawing the network, one might include only ties between people and their friends and spouses while excluding ties to siblings and co-workers. Or, in a network that contained people's connections, one might include only individuals with whom people had had sex or transacted business. Moreover, one might look at just the largest component of a network, or one might sample several hundred nodes from the network to study part of its structure more closely.

It is important to note, however, that while the fundamental pattern of ties in a social network (its topology) is fixed, how this pattern is visually rendered in two-dimensional space depends on the analyst's objectives. Imagine a set of 500 buttons strewn on the floor. And imagine that there are 2,000 strings we can use to connect the buttons. We pick two buttons at random and connect them with a string. Then we repeat this procedure until all the strings are used up. Some buttons will have many strings attached to them (and hence have high degree), and others, by chance, will never have been picked and so will not be connected to any other button. Perhaps some groups of buttons will be connected to each other but be separated from other groups because no string connects any button in one group to any button in the other group. These would be two components of the network. If we were to pick up one button in one component and lift it up from the floor, all the other buttons it was attached to, directly or indirectly, would follow it up into the air as we lifted it. And if we were to drop the mass of buttons onto another spot on the floor, it would look different than when we picked it up. But the fundamental topology would be exactly the same.

The challenge in network visualization procedures is to specify a way of showing this fundamental and invariant topology in an appealing and faithful way. Several procedures are available to draw networks in two-dimensional space. Most procedures rely on so-called "spring embedder" mechanisms, which see the ties as springs and which iteratively reposition nodes so as to minimize the total "energy" in the system. The widely used Kamada-Kawai algorithm,[xiv] implemented in Pajek software,[xv] iteratively repositions nodes in order to reduce the number of ties that cross each other, while also maintaining a certain distance between the nodes to minimize their overlap. It generates a matrix of shortest network path distances from each node to all other nodes in the network and repositions nodes so as to reduce the sum of the difference between the plotted distances and the network distances. The Fruchterman Rheingold algorithm,[xvi] also implemented in Pajek, gives a somewhat different visualization, as it tends to place nodes in tightly-knit communities much closer together and much further from other parts of the network.

There are a number of social network data sets that contain information about nodes and ties that are longitudinal, namely, that record information across time. Nodes might come and go (people are born or die), ties might form and break (people make new friends) or individuals might change (they might gain or lose weight, for example). Visualizing such temporally evolving networks can involve other software such as SONIA,[xvii] and examples of the videos such software can generate are available at the author's websites. Examples of such longitudinal network datasets include the National Longitudinal Study of Adolescent Health (Add Health),[xviii] and the Framingham Heart Study Social Network (FHS-Net). Epidemiological studies of sexually transmitted disease networks often also have, or benefit from, such longitudinal data about who is having sex with whom.[xix]

Statistical Analysis of Networks

Individuals connected to each other in a network may evince similar characteristics (such as having similar academic performance, body size, health behaviors, or political views), and clustering within the network based on such characteristics can be visualized in network graphs. Such clustering, however, could be attributed to at least three processes: 1) *homophily*, whereby individuals choose to become connected based on shared attributes or behaviors (i.e., the tendency of like to attract like, or of "birds of a feather to flock together");[xx] 2) *induction,* whereby attributes or behaviors in one person cause analogous attributes or behaviors in others; or 3) *confounding*, whereby connected individuals jointly experience contemporaneous exposures (that cause them both to share an attribute or behavior). To distinguish among these effects using observational data requires repeated measures of the attributes,[xxi] longitudinal information about network ties, and information about the nature or direction of the ties (e.g., who nominated whom as a friend).

The statistical analysis of networks can focus on the processes responsible for network structure (such as homophily), or the processes involved in network function (such as induction). With respect to the analysis of processes determining the structure of the network, one might ask such questions as: Why do networks have a particular structure? What determines which ties exist? Such processes might include homophily, rules governing a predilection to transitivity, or other rules governing the link-forming process (such as preferential attachment to high-degree individuals, or even genetic heterogeneity in individuals' taste for network location [xxii]). Such analyses may use statistical models such as so-called p* models or conventional regression analysis (especially in sparse networks where the number of ties is low compared to the number of nodes).[xxiii,xxiv]

The statistical analysis of networks may also focus on the processes by which attributes or behaviors spread across the network. These two analytic frames can overlap. Some statistical packages attempt to model both processes simultaneously, such as SIENA, but this software is presently limited in the size of the networks it can handle.[xxv] Other approaches use more conventional regression techniques, with adjustments required by the application to networks (such as the non-independence of the observations).[xxvi] Tools for analyzing the spread of phenomena in networks across time when the network is itself evolving are still incomplete and under active statistical development. For example, the directionality of arcs may be exploited to support causal inference.[xxvii,xxviii,xxix]

There are a variety of other special statistical issues involved in the study of networks. For example data are often missing, and, in a network situation, they can be missing in quite a number of complex ways; nodes, ties, attributes, or observation waves may be missing, and this missingness affects not just the individual observation, but others to which it is connected. Furthermore, the network under study may be only partially observed (either by design, as in sampling, or unavoidably, as in observational studies), and people within the network may have ties to others who are not observed by the investigator; when this process is associated with attributes of the people being studied, this can lead to confounding. The current best practice for resolving this problem is to compare the distribution of the dependent variable between people who have ties inside the network and those who have ties outside the network. If the two distributions are not statistically different, then it suggests missingness will not have a direct impact on the estimates of associations within the group that is fully observed.

The Framingham Heart Study Social Network

The Framingham Heart Study (FHS), a landmark epidemiological study initiated in 1948, has yielded important findings about cardiovascular risk factors and other health phenomena since its inception. When it was initiated in 1948, 5,209 people in Framingham, Massachusetts were enrolled into the "Original Cohort." In 1971, the "Offspring Cohort," composed of most of the children of the Original Cohort, and their spouses, was enrolled. This cohort of 5,124 people has had almost no loss to follow-up other than death (only 10 cases dropped out). In 2002, enrollment of the so-called "Third Generation Cohort" began, consisting of 4,095 of the children of the Offspring Cohort. The Framingham Heart Study also involves certain other smaller cohorts (*e.g.*, a minority over-sample called the OMNI Cohort enrolled in 1995). Participants in all these cohorts come to a central facility for detailed examinations and survey data collection at regular intervals. For example, there have been seven waves of data collection since 1971 in the Offspring cohort, roughly every four years.[xxx]

To develop a new dataset based on the FHS, which we call the FHS-Net, we used the Offspring Cohort as the source of 5,124 egos to study. Each ego in this cohort is connected to other people (the alters) via friendship, family, spousal, neighbour, and coworker relationships. Overall, within the entire FHS social network composed of both the egos and any detected alters in any FHS cohort, there are 12,067 individuals who were connected at some point during the period 1971 to 2003.

To create the network dataset, we computerized information about the Offspring Cohort from archived, handwritten administrative tracking sheets that had been used since 1971 to identify people close to participants for the purpose of facilitating follow-up. These documents contain valuable, previously unused social network information because subjects were asked to identify their relatives, "close friends," place of residence, and place of work in order to ensure they could be contacted every two to four years. In the field of network science, such procedures for identifying social ties between individuals are known as "name generators."[xxxi,xxxii]

Moreover, this dataset identifies the network links among participants longitudinally at each wave. Over the course of follow-up, the participants spread out across the USA, but they nevertheless continued to participate in the FHS. As a person's family changed due to birth, death, marriage, or divorce, and as their contacts changed due to residential moves, new places of employment, or new friendships, this information was captured.

Overall, there were 53,228 observed social ties between the 5,124 egos and any alters in any of the FHS cohorts, yielding an average of 10.4 ties to family, friends, and co-workers over the course of follow-up. Additional ties to neighbours were also ascertained, based on information about all participants' place of residence, but they are not included in the foregoing count since the number of neighbour ties depends on how "neighbour" is defined (e.g., whether we restrict the definition to immediate, "next-door" neighbours, or neighbours residing on the same block within 25 or 100 meters, etc.). For example, one ego in the Offspring Cohort had 18 alters: a mother, a father, a sister, two brothers, three children, two friends, five neighbours (living within 25 meters), and three coworkers, and all these individuals were themselves in the network and repeatedly observed since 1971.

Given the compact nature of the Framingham social network in the period since 1971, many of the nominated contacts were, as noted, themselves also participants of one or another FHS cohort – which is a crucial feature. This means that detailed survey and physical exam information about both the ego and the alters is available. For example, 83% of egos' spouses were directly and repeatedly observed and 87% of egos with siblings had at least

one sibling who also participated in the FHS. For 39% of the egos, at least one co-worker participated in the study. For 10% of the egos, an immediate neighbour was also in the FHS (more expansive definitions of neighbours, such as those people living within 100 meters, yielded much higher percentages).

Importantly, 45% of the 5,124 egos were connected via friendship to another person in the study; there were 3,604 unique, observed friendships for an average of 0.7 friendship ties per ego. However, there was substantial variation from person to person, ranging from several people with no friends to one person who was nominated as a friend by eight different FHS participants. Because friendship identifications are directional arcs, we can study three different types. An “ego-perceived friend” means the ego nominates an alter as a friend, but the friendship nomination is not reciprocated. In this case the ego thinks of the alter as a friend, but the alter *may not* think of the alter as a friend. An “alter-perceived friend” means the alter nominates the ego as a friend but not vice versa. Here, the ego may not feel any closer to the alter than he or she would to a stranger. Finally, a “mutual friend” is one in which the nomination is reciprocal. This directional information can be exploited for causal inference in social networks, helping to address issues of endogeneity and confounding. [xxxiii,xxxiv,xxxv]

We evaluated whether the FHS-Net resembled a small-world,[xxxvi] scale-free,[xxxvii] or hierarchical network.[xxxviii] Examination of the degree distribution in the FHS-Net revealed that most individuals have one or two close friends and 10 or fewer family members who also participate in the Framingham Study (the number of close friends is in keeping with other national studies [xxxix,xl]). A small number of nodes are very well-connected, as also observed in other social networks. However, the degree distribution suggests that the FHS-Net does not conform either to a small world network model nor to the scale-free model.[xli]

Findings in the FHS Social Network

Our study of the FHS-Net has documented the clustering of individuals with similar characteristics within the network. In particular, we have found that obesity, smoking behavior, and happiness show clustering.[xlii,xliii,xliv] Importantly, this clustering is not solely due to homophily, and our analyses provide diverse sorts of evidence for the likelihood of spread of these traits, that is, evidence of induction. In addition, we find that these clusters extend to three degrees of separation, such that, from the point of view of an ego, their weight status, smoking behavior, and happiness are related to the analogous attributes of individuals three degrees removed from them (e.g, their friends’ friends’ friends). These findings provide support for the idea that these behaviors and states have collective and not just individual properties and determinants.

Here, we show three figures illustrating various aspects of our findings. Figure 1 shows the largest connected sub-component of the FHS-Net in the year 2000, with graphical features highlighting obesity. Specifically, node size is made proportional to people’s body mass index (BMI), and nodes are colored yellow if the BMI is above 30 (which indicates obesity). This network is sufficiently dense to obscure much of the underlying structure, though regions of the network with clusters of obese or non-obese persons can be observed.

Figure 2 shows part of the FNS-Net in 1971 and 2000. Again, node size is made proportional to an individual’s attribute, here, the number of cigarettes smoked, and nodes are colored yellow if a person smoked more than one cigarette per day. There is substantial change in the prevalence of smoking and in the social life of smokers. In 1971, there were many more smokers as compared to 2000, and the smokers occupied the center of their

circles of friends and family to the same extent that non-smokers did. However, by 2000, most people had quit smoking, and those who still smoked were more likely to be at the periphery of the network, which is visible by inspection (and confirmed statistically). Moreover, there was an increased tendency for smokers to be connected primarily to other smokers and for there to be relatively separate clusters of smokers and non-smokers. And whole clusters of smokers quit smoking together.

Figure 3 shows the largest connected network component in 2000 based on a restricted set of ties among siblings, spouses, and friends (co-workers and neighbours are again excluded to simplify the image), and the objective here was to study the role of social networks in emotional states like happiness (namely, whether emotions have a collective, and not merely individual, origin). To highlight the clustering of happiness, which is apparent upon visual inspection, each node is colored according to the subject's happiness on a spectrum from blue (unhappy) to yellow (happy). In this particular visualization we used a technique that we call "social space smoothing." To better identify large-scale structure in the network, we recoded the happiness of each person to be equal to the average of the individual's happiness and that of all people to whom he or she is connected (i.e. the ego and all his or her first-degree alters). Visual inspection of Figure 3 shows that there are large social "neighborhoods" that tend to be more happy and others that tend to by less happy. Figure 3 also suggests a relationship between network centrality and happiness: subjects at the core of their local networks appear more likely to be happy, while those on the periphery appear more likely to be unhappy.

All of the foregoing visualizations were coupled with statistical analyses of the kinds described earlier. In ongoing work, we are investigating how alcohol consumption, eating behavior, depression, loneliness, and health screening behavior might spread in social networks. Other contemporary phenomena that might evince epidemic properties include peanut allergies [xlv] and autism diagnoses (as some work by Peter Bearman may suggest); both of these may have a social network component in that as individuals note others around them with these conditions, they may be more prone to being diagnosed with them themselves.

Online Networks

Many investigators have been examining online social networks with respect to health-related phenomena and are beginning to use online networks as opportunities for interventions.[xlvi,xlvii] One of our efforts has involved the examination of a group of 1,700 college students who are interconnected in Facebook.[xlviii] We examined these students' online profiles, noting their friends and their appearance in photographs.

The photographs were valuable in two ways. First, we coded who appeared in photographs with whom. People who take the trouble to be in the same place, take a photograph together, upload the photograph, and label ("tag") it, almost certainly have a closer relationship with one another than the usual "friends" people indicate in online social networking sites. In fact, while the average student in our data had over 110 friends on Facebook, they had an average of only six "picture friends" (i.e., people close enough that they tagged the student). Second, we coded whether the students were smiling in their profile photographs (as well as other physical attributes), and we mapped the network of students and their picture friends, making note of who was smiling and who was not.

Figure 4 is a map of part of this Facebook network in 2007. It contains 353 students; the lines between nodes indicate that the connected individuals were tagged in a photo together. Once again we used "social space smoothing" to highlight structure in the network. Students who are smiling (and who are immediately surrounded by smiling people in their network)

are colored yellow. Students who are frowning (and who are immediately surrounded by such serious looks) are colored blue. As in the Framingham happiness study, here the blue nodes and the yellow nodes cluster together, indicating large-scale structure of smiling "neighborhoods" in the online network. Visual inspection suggests, and statistical analyses confirm, that those who smile are measurably more central to the network compared to those who do not smile. Moreover, statistical analysis of the network shows that people who smile tend to have more friends (smiling is associated with having an average of one extra friend, which is impressive considering that people only have about six close friends).

Practical Utility of Network Graphs

At present, there is much interest in using networks as means to spread positive health behaviors or as targets of health interventions. This impetus is supported by prior work on peer effects. For example, the smoking behavior of an adolescent's friends influences the odds of smoking initiation, continuation, and cessation.[xlix,l,li] Similar effects are seen in alcohol use.[lii,liii] Smoking and alcohol cessation programs that provide peer support – that is, that modify the social network of the target – have been shown to be more successful. [liv,lv,lvi] Like tobacco and alcohol consumption, behaviors related to weight also appear to be socially transmissible. Studies have linked unhealthy weight-control behaviors among adolescent girls to the dieting behaviors of their peers [lvii], and children's food preferences have been shown to be manipulable using peer modeling [lviii]. Among adults, delivering a successful weight loss intervention to one person has been shown to trigger substantial weight loss in that person's friends, and there is evidence to suggest that weight loss interventions that target social networks are more effective than are those that target isolated individuals.[lix,lx,lxi,lxii] Weight loss has also been shown to spread across social ties in intervention trials as well.[lxiii]

Network visualizations can support such interventions in numerous ways (either with or without supplementary statistical analyses). First, they can be used to identify cliques or clusters of individuals with similar health-relevant attributes within which reinforcement of positive or negative behaviors may be taking place. These cliques could be targeted for collective interventions.

Second, they can be used to identify and target individuals for public health interventions. Peripheral individuals might be at risk for ill health by virtue of social isolation, and central ones might be appealing targets for intervention since they may be particularly influential. [lxiv,lxv]

Third, a knowledge of the overall network structure may be crucial to the design of public health interventions. A compelling example is provided by the case of sexually transmitted diseases (STDs). Bearman and Moody used a sub-sample of Add Health data to model the complete sexual network of a mid-sized, predominantly white Midwestern high school using information on reported romantic partnerships over an 18-month period.[lxvi] They found that a surprisingly sizeable 52% of all romantically-involved students were embedded in one very large ring network with branches, which they described as a "spanning tree." This spanning tree was especially notable for its lack of redundant ties, meaning that most students were connected to the superstructure by one pathway only.

Most models of STD transmission assume the existence of high activity "cores" that disseminate disease to lower activity individuals and that sustain epidemics by functioning as reservoirs of infection. As Bearman and Moody point out, however, their findings are significant both for their inconsistency with this traditional notion of core groups as the drivers of STD diffusion and for their implications for STD control, which stem from the largest component's fragility: if a link from the "trunk" of the spanning tree is removed, the

transmission of infection beyond that linkage is effectively halted as the network breaks into two disjoint components. As such, the network they documented was highly vulnerable to the removal of single ties or nodes, which, they argue, is best achieved by broad-based, "broadcast" STD control programs – that is, those that target the entire population rather than specific activity groups.

In studying similar dynamics with respect to the HIV/AIDS epidemic in Sub-Saharan Africa, Helleringer and Kohler collected information on up to five recent sexual partners of the residents of seven villages located on an island in Lake Malawi.[lxvii] They found that, contrary to expectations, residents reported relatively few partners. Despite this finding, upon mapping the resulting sexual network, they discovered that a striking 65% of the population aged 18–35 was connected in one large interconnected component. However, unlike in the Bearman and Moody study, this large component was strikingly robust to the removal of individual ties or nodes as a result of numerous redundant paths (i.e., instances in which respondents directly or indirectly shared more than one sexual partner). Yet, like Bearman and Moody, Helleringer and Kohler failed to find evidence of high activity hubs, that is, persons or groups capable of sustaining the HIV/AIDS epidemic by having many sexual partners. As they note, their findings thus call into question the assumption behind much HIV work in Sub-Saharan Africa, that the current epidemic is driven either by a high activity core made up of sex workers and their patrons or by other high activity individuals transmitting disease to a low activity "periphery" made up of individuals with one or few partners.

In addition to the insights they provide into mechanisms underlying the spread of STDs and, consequently, methods for possible containment, these studies are important for demonstrating the value of collecting sociocentric network data (as opposed to egocentric network data collected from a series of not necessarily inter-connected individuals). Without sociocentric data, the contact macrostructure through which infectious disease – or, alternatively, influence, information, or other socially transmissible constructs – must flow could not be mapped and understood.

Sources of Network Data

Collecting network data is not easy, especially if a full sociocentric study is done and especially if the intention is that it be longitudinal. Therefore, people are beginning to explore the use of readily available or extant data, such as that available using social network websites such as facebook or MySpace [lxviii,lxix] or mobile phone networks [lxx] or perhaps overlapping lists of memberships in organizations or clubs or use of public services. Such sources of data would allow us to visualize enormous networks composed of hundreds of thousands of people and to intervene on them as well.

Ethical problems will inevitably arise with using such network data, particularly with respect to concerns about anonymity. However, a more thoroughgoing ethical issue is suggested by the very nature of network health effects: more connected individuals may be more valuable or more worthy of receiving medical care, given the effect they have on others.[lxxi] Moreover, health care delivered to the well connected is clearly more cost-effective since the effect such individuals have on others results in more "Quality Adjusted Life Years" per dollar spent.[lxxii,lxxiii]

Provider Networks

One can use similar methods to those we have discussed in order to visualize networks of health care providers, such as doctors in a community who consult each other or who share patients,[lxxiv] or hospitals that transfer patients to each other. Such networks have

important implications. For example, physicians' locations within networks of colleagues may serve to make some physicians aware of innovations in medicine sooner than others.[lxxv] Local opinion leaders occupying strategic, central network positions may disseminate influential assessments of both established and innovative medical regimens.[lxxvi] Physicians may also look to nearby role models within their social networks for guidance in treating their patients.[lxxvii,lxxviii] Social networks therefore can shape health care delivery.

Conclusion

People are interconnected and so their health is interconnected. The recognition that people are embedded in social networks means that the health and well being of one person affects the health and wellbeing of others. This fundamental fact of existence provides a fundamental conceptual justification for the field of public health. Visualizing social networks and the health-relevant phenomena that transpire within them provides a new way to understand the epidemiological determinants of illness and wellbeing.

Acknowledgments

Supported by NIH/NIA P-01 AG-031093.

References

i. Berkman LF, Leo-Summer L, Horwitz RI. Emotional Support and Survival After Myocardial Infarction: A Prospective, Population-based Study of the Elderly. Annals of Internal Medicine. 1992; 117:1003–1009. [PubMed: 1443968]

ii. House JS, Landis KR, Umberson D. Social Relationships and Health. Science. 1988; 241:540–545. [PubMed: 3399889]

iii. Cohen S, Doyle WJ, Skoner DP, Rabin BS, Gwaltney JM Jr. Social ties and susceptibility to the common cold. JAMA. 1997; 277:1940–1944. [PubMed: 9200634]

iv. Laumann EO, Youm Y. Racial/Ethnic Group Differences in the Prevalence of Sexually Transmitted Diseases in the United States: a Network Explanation. Sexually Transmitted Diseases. 1999; 26:250–61. [PubMed: 10333277]

v. Liljeros F, Edling CR, Amaral LAN, Stanley HE, Aberg Y. The web of human sexual contacts. Nature. 2001; 411:907–908. [PubMed: 11418846]

vi. Potterat JJ, Phillips-Plummer L, Muth SQ, Rothenberg RB, Woodhouse DE, Maldonado-Long TS, Zimmerman HP, Muth JB. Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs. Sex Transm Infect. 2002; 78(90001):159i–163.

vii. Berkman, L.; Glass, T. Social Integration, Social Networks, and Health. In: Berkman, LF.; Kawachi, I., editors. Social Epidemiology. Oxford: Oxford University Press; 2000. p. 174-190.

viii. Bonacich P. Factoring and Weighing Approaches to Clique Identification. Journal of Mathematical Sociology. 1972; 2:113–120.

ix. Freeman LC. Centrality in Social Networks. I. Conceptual Clarification. Social Networks. 1979; 1:215–39.

x. Watts DJ, Strogatz SH. Collective Dynamics of Small World Networks. Nature. 1998; 393:490–410.

xi. Watts DJ. Networks, Dynamics, and the Small-World Phenomenon. American Journal of Sociology. 1999; 105:493–527.

xii. Barabasi A-L, Albert R. Emergence of scaling in random networks. Science. 1999; 286:509–512. [PubMed: 10521342]

xiii. Fowler JH, Dawes CT, Christakis NA. Model of Genetic Variation in Human Social Networks. PNAS. 2009 in press.

xiv. Kamada T, Kawai S. An algorithm for drawing general undirected graphs. Information Processing Letters. 1989; 31:7–15.

xv. Batagelj, V.; Mrvar, A. Program for Analysis and Visualization of Large Networks, version 1.14. 2006.

xvi. Fruchterman TMJ, Reingold EM. Graph Drawing by Force-Directed Placement. Software: Practice and Experience. 1991; 21(11)

xvii. Moody J, McFarland DA, Bender-deMoll S. Visualizing Network Dynamics. American Journal of Sociology. 2005; 110:1206–41.

xviii. Udry, JR. The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002 [machine-readable data file and documentation]. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill; 2003.

xix. Morris M, Kretzschmar M. Concurrent partnerships and the spread of HIV. AIDS. 1997; 11:641–648. [PubMed: 9108946]

xx. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Annual Review of Sociology. 2001; 27:415–444.

xxi. Carrington, PJ.; Scott, J.; Wasserman, S. Models and Methods in Social Network Analysis. Cambridge: Cambridge University Press; 2005.

xxii. Fowler JH, Dawes CT, Christakis NA. Model of Genetic Variation in Human Social Networks. PNAS. 2009 in press.

xxiii. OMalley AJ, Marsden PV. The Analysis of Social Networks. Health Services and Outcomes Research Methodology. 2008; 8:222–269. [PubMed: 20046802]

xxiv. OMalley AJ, Christakis NA. Health Behaviors and Traits and the Formation and Dissolution of Friendship Ties. Social Networks. forthcoming.

xxv. Snijders, T.; Steglich, C.; Schweinberger, M.; Huisman, M. Manual for SIENA Version 3.2. University of Groningen; Groningen, The Netherlands: 2007.

xxvi. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med. 2007; 357:370–379. [PubMed: 17652652]

xxvii. Fowler JH, Christakis NA. Estimating peer effects on health in social networks. Journal of Health Economics. 2008; 27:1400–1405. [PubMed: 18692263]

xxviii. Bramoullé, Y.; Djebbari, H.; Fortin, B. IZA Discussion Papers. Institute for the Study of Labor (IZA); 2007. Identification of Peer Effects through Social Networks; p. 2652

xxix. Anagnostopoulos, A.; Kumar, R.; Mahdian, M. Influence and correlation in social networks. KDD ‘08; Proceeding of the 14th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining; Las Vegas, Nevada, USA. August 24–27, 2008; New York, NY: ACM; 2008. p. 7-15.DOI= http://doi.acm.org/10.1145/1401890.1401897

xxx. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med. 2007; 357:370–379. [PubMed: 17652652]

xxxi. Campbell KE, Lee BA. Name Generators In Surveys Of Personal Networks. Social Networks. 1991; 13:203–221.

xxxii. Marsden PV. Interviewer Effects In Measuring Network Size Using A Single Name Generator. Social Networks. 2003; 25:1–16.

xxxiii. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med. 2007; 357:370–379. [PubMed: 17652652]

xxxiv. Bramoullé, Y.; Djebbari, H.; Fortin, B. IZA Discussion Papers. Institute for the Study of Labor (IZA); 2007. Identification of Peer Effects through Social Networks; p. 2652

xxxv. Anagnostopoulos, A.; Kumar, R.; Mahdian, M. Influence and correlation in social networks. KDD ‘08; Proceeding of the 14th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining; Las Vegas, Nevada, USA. August 24–27, 2008; New York, NY: ACM; 2008. p. 7-15.DOI= http://doi.acm.org/10.1145/1401890.1401897

xxxvi. Watts DJ, Strogatz SH. Collective Dynamics of Small World Networks. Nature. 1998; 393:490–410.

xxxvii. Barabasi AL, Albert R. Emergence of scaling in random networks. Science. 1999; 286:509–512. [PubMed: 10521342]

xxxviii. Ravasz E, Barabasi AL. Hierarchical organization in complex networks. Physical Review E. 2003; 67:026112.

xxxix. Marsden PV. Core Discussion Networks of Americans. American Sociological Review. 1987; 52:122–131.

xl. McPherson M, Smith-Lovin L, Brashears ME. Social isolation in America: changes in core discussion networks over two decades. American Sociological Review. 2006; 71:353–375.

xli. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med. 2007; 357:370–379. [PubMed: 17652652]

xlii. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med. 2007; 357:370–379. [PubMed: 17652652]

xliii. Christakis NA, Fowler JH. The Collective Dynamics of Smoking in a Large Social Network. N Engl J Med. 2008; 358:2249–2258. [PubMed: 18499567]

xliv. Fowler JH, Christakis NA. Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 Years in the Framingham Heart Study. British Medical Journal. 2008; 337:a2338. [PubMed: 19056788]

xlv. Christakis NA. This Allergies Hysteria Is Just Nuts. British Medical Journal. 2008; 337:1384.

xlvi. Moreno MA, Parks MR, Zimmerman FJ, Brito TE, Christakis DA. Display of health risk behaviors on MySpace by adolescents. Arch Pediatr Adolesc Med. 2009; 163:27–34. [PubMed: 19124700]

xlvii. Moreno MA, VanderStoep A, Parks MR, Zimmerman FJ, Kurth A, Christakis DA. Reducing at-risk adolescents' display or risk behaviors on a social networking web site: a randomized controlled pilot intervention trial. Arch Pediatr Adolesc Med. 2009; 163:35–41. [PubMed: 19124701]

xlviii. Lewis K, Kaufman J, Gonzalez M, Wimmer A, Christakis NA. Tastes, Ties, and Time: A New (Cultural, Multiplex, and Longitudinal) Social Network Dataset Using Facebook. com Social Networks. 2008; 30:330–342.

xlix. Burt RD, Peterson AV Jr. Smoking cessation among high school seniors. Preventive Medicine. 1998; 27:319–27. [PubMed: 9612822]

l. Kaplan CP, Napoles-Springer A, Stewart SL, Perez-Stable EJ. Smoking acquisition among adolescents and young Latinas: the role of socioenvironmental and personal factors. Addictive Behaviors. 2001; 26:531–50. [PubMed: 11456076]

li. Chen PH, White HR, Pandina RJ. Predictors of smoking cessation from adolescence into young adulthood. Addictive Behaviors. 2001; 26:517–29. [PubMed: 11456075]

lii. Andrews JA, Tildesley E, Hops H, Li F. The influence of peers on young adult substance use. Health Psychology. 2002; 21:349–57. [PubMed: 12090677]

liii. Urberg KA, Degirmencioglu SM, Pilgrim C. Close friend and group influence on adolescent cigarette smoking and alcohol use. Developmental Psychology. 1997; 33:834–44. [PubMed: 9300216]

liv. Prince F. The relative effectiveness of a peer-led and adult-led smoking intervention program. Adolescence. 1995; 30:187–94. [PubMed: 7625255]

lv. Malchodi CS, Oncken C, Dornelas EA, Caramanica L, Gregonis E, Curry SL. The effects of peer counseling on smoking cessation and reduction. Obstetrics & Gynecology. 2003; 101:504–10. [PubMed: 12636954]

lvi. McKnight AJ, McPherson K. Evaluation of peer intervention training for high school alcohol safety education. Accident Analysis & Prevention. 1986; 18:339–47. [PubMed: 3741584]

lvii. Eisenberg ME, Neumark-Sztainer D, Story M, Perry C. The Role of Social Norms and Friends' Influences on Unhealthy Weight-Control Behaviors Among Adolescent Girls. Social Science and Medicine. 2005; 60:1165–1173. [PubMed: 15626514]

lviii. Lowe CF, Horne PJ, Tapper K, Bowdery M, Egerton C. Effects of a Peer Modelling and Rewards-Based Intervention to Increase Fruit and Vegetable Consumption in Children. European Journal of Clinical Nutrition. 2004; 58:510–522. [PubMed: 14985691]

lix. Wing RR, Jeffery RW. Benefits of Recruiting Participants With Friends and Increasing Social Support for Weight Loss and Maintenance. Journal of Consulting and Clinical Psychology. 1999; 67:132–8. [PubMed: 10028217]

lx. Verheijden MW, Bakx JC, Van Weel C, Koelen MA, Van Staveren WA. Role of Social Support in Lifestyle-Focused Weight Management Interventions. Eur J Clin Nutr. 2005; 59:S179–S186. [PubMed: 16052189]

lxi. Gorin A, Phelan S, Tate D, Sherwood N, Jeffery R, Wing R. Involving Support Partners in Obesity Treatment. Journal of Consulting and Clinical Psychology. 2005; 73:341–3. [PubMed: 15796642]

lxii. Kelsey K, Earp JL, Kirkley BG. Is Social Support Beneficial for Dietary Change? A Review of the Literature Family and Community Health. 1997; 20:70–82.

lxiii. Gorin AA, Wing RR, Fava JL, et al. International Journal of Obesity. 2008; 32:1678–1684. [PubMed: 18762804]

lxiv. Valente TW, Hoffman BR, Ritt-Olson A, Lichtman K, Johnson CA. Effects of a social-network method for group assignment strategies on peer-led tobacco prevention programs in schools. Am J Publ Health. 2003; 93:1837–1843.

lxv. Kelly JA, Amirkhanian YA, Kabakchieva E, et al. Prevention of HIV and sexually transmitted diseases in high risk social networks of young Roma (Gypsy) men in Bulgaria: randomized controlled trial. British Medical Journal. 2006; 333:1098–1101. [PubMed: 17040924]

lxvi. Bearman PS, Moody J, Stovel K. Chains of Affection: the Structure of Adolescent Romantic and Sexual Networks. American Journal of Sociology. 2004; 110:44–91.

lxvii. Helleringer S, Kohler H-P. Sexual Network Structure and the Spread of HIV in Africa: Evidence from Likoma Island, Malawi. AIDS. 2007; 21:2323–2332. [PubMed: 18090281]

lxviii. Lewis K, Kaufman J, Gonzalez M, Wimmer A, Christakis NA. Tastes, Ties, and Time: A New (Cultural, Multiplex, and Longitudinal) Social Network Dataset Using Facebook. com Social Networks. 2008; 30:330–342.

lxix. Moreno MA, Parks MR, Zimmerman FJ, Brito TE, Christakis DA. Display of health risk behaviors on MySpace by adolescents. Arch Pediatr Adolesc Med. 2009; 163:27–34. [PubMed: 19124700]

lxx. Onnela JP, Saramaki J, Hyvonen J, Szabo G, Lazar D, Kaski K, Kertesz J, Barabasi AL. Structure and Tie Strengths in mobile communication networks. PNAS. 2007; 104:7332–7336. [PubMed: 17456605]

lxxi. Christakis NA. Valuing the Well Connected. British Medical Journal. 2008; 337:725.

lxxii. Christakis NA, Allison PD. Mortality after the hospitalization of a spouse. New England Journal of Medicine. 2006; 354:719–730. [PubMed: 16481639]

lxxiii. Christakis NA. Social networks and collateral health effects – have been ignored in medical care and clinical trials, but need to be studied. British Medical Journal. 2004; 329:184–185. [PubMed: 15271805]

lxxiv. Keating NL, Ayanian JZ, Cleary PD, Marsden PV. Factors affecting influential discussions among physicians: a social network analysis of primary care practice. Journal of General Internal Medicine. 2007; 22:794–798. [PubMed: 17404798]

lxxv. Coleman, JS.; Katz, E.; Menzel, H. Medical Innovation: A Diffusion Study. Indianapolis, IN: Bobbs-Merrill; 1966.

lxxvi. Weimann, G. The Influentials: People Who Influence People. Albany, NY: State University of New York Press; 1994.

lxxvii. Burt RS. Social Contagion and Innovation: Cohesion versus Structural Equivalence. American Journal of Sociology. 1987; 92:1287–335.

lxxviii. Keating NL, Ayanian JZ, Cleary PD, Marsden PV. Factors affecting influential discussions among physicians: a social network analysis of primary care practice. Journal of General Internal Medicine. 2007; 22:794–798. [PubMed: 17404798]

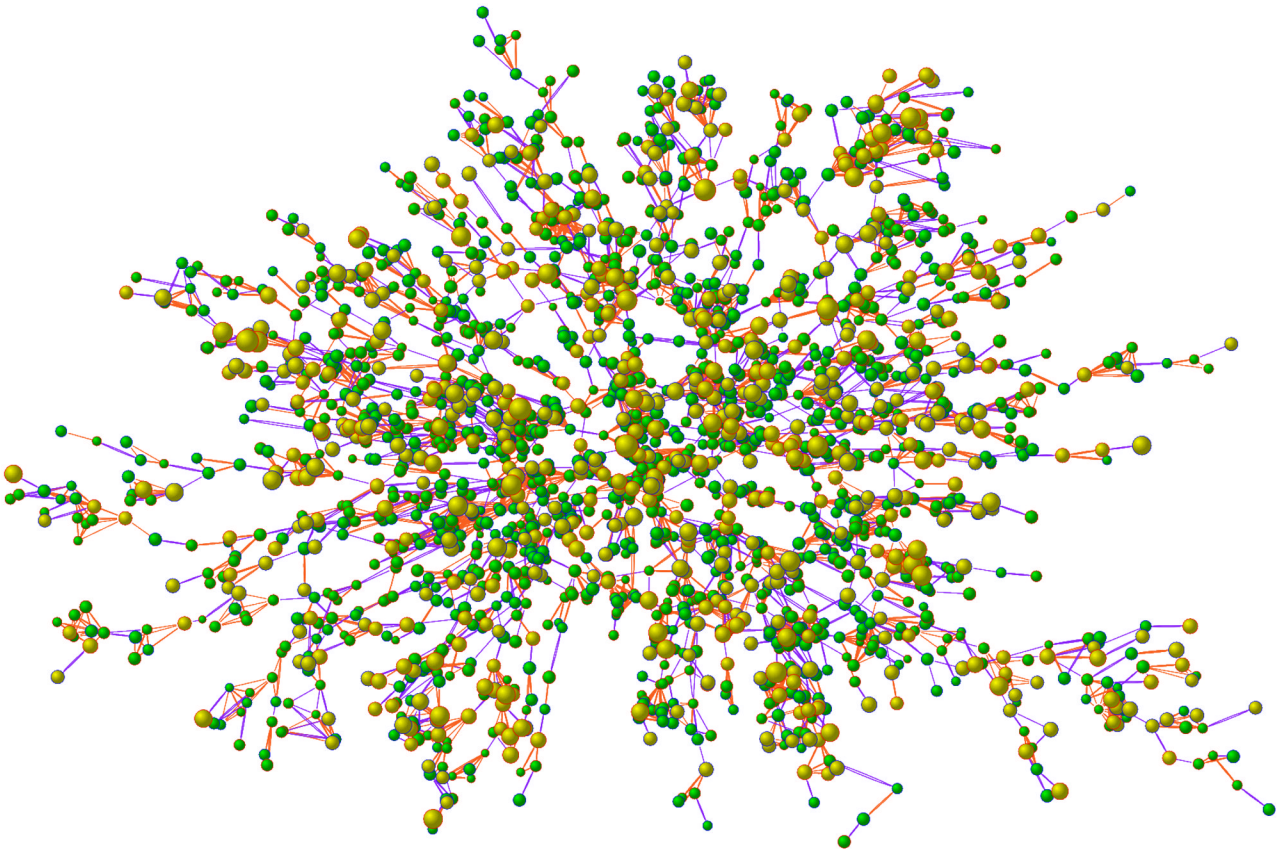

Figure 1. Obesity Clusters in the Framingham Social Network

This is the largest connected component of the Framingham Heart Study Social Network in 2000 (N=2,200). Node border indicates gender (red=female subject, blue=male subject), node color indicates obesity (yellow=BMI>30), node size is proportional to BMI, and tie colors indicate relationship (purple=friend or spouse, orange=family). Clusters of obese and non-obese individuals are visible, though the complexity of the image is still very high.

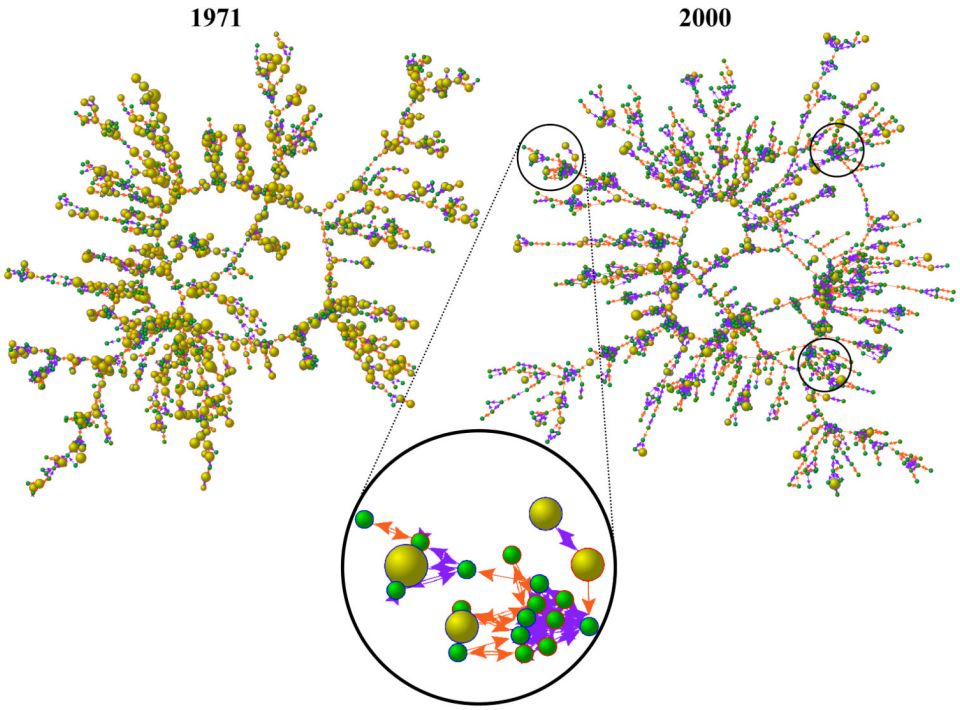

Figure 2. Smoking in the Framingham Social Network

This is a random sample of 1000 subjects in the FHS social network chosen from the largest connected subcomponent at exam 1 (left) and exam 7 (right). Node border indicates gender (red=female, blue=male), node color indicates cigarette consumption (yellow is for ≥1 cigarettes per day), node size is proportional to number of cigarettes consumed, and arrow colors indicate relationship (friends and spouses = orange, family = purple). By 2000, it is apparent that smokers are more likely to occur at the periphery of their networks. And smokers are usually in smaller subgroups than nonsmokers. The circles in the panel for 2000 identify densely connected clusters of green circles where there are no smokers at all or where the smokers sit at the periphery of the subgroup.

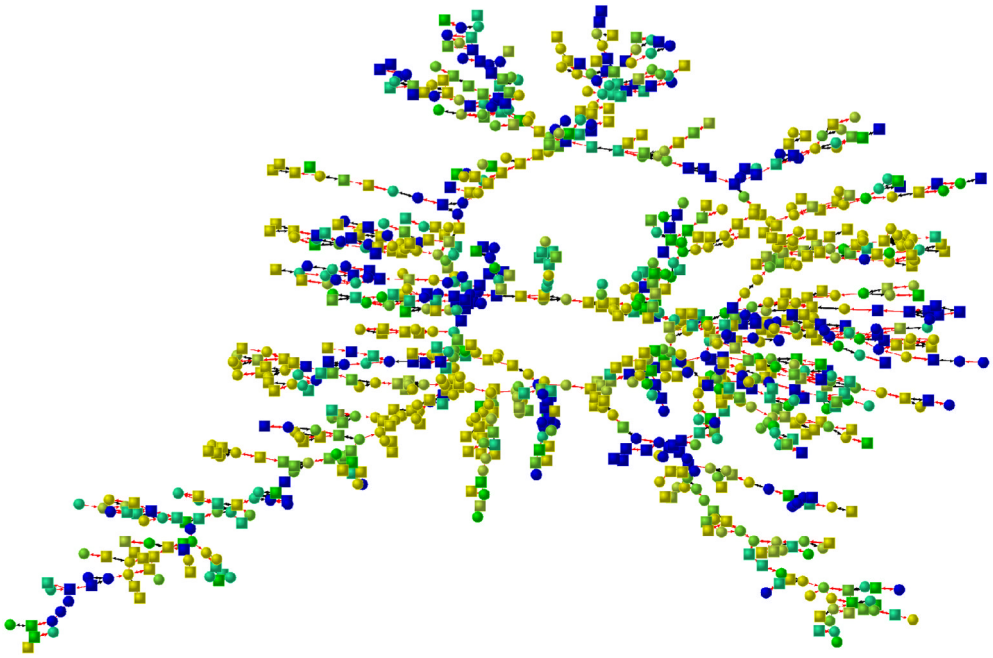

Figure 3. Happiness Clusters in the Framingham Social Network

This graph shows the largest component of friends, spouses, and siblings in the year 2000. There are 1020 individuals shown. Each node represents a subject and its shape denotes gender (circles are female, squares are male). Lines between nodes indicate relationship (black for siblings, red for friends and spouses). Node color denotes the mean happiness of the ego and all directly connected (distance 1) alters, with blue shades indicating the least happy, and yellow shades indicating the most happy (shades of green are intermediate). Clusters of happy and unhappy individuals are visible, and unhappy individuals appear more likely to be at the periphery of the network.

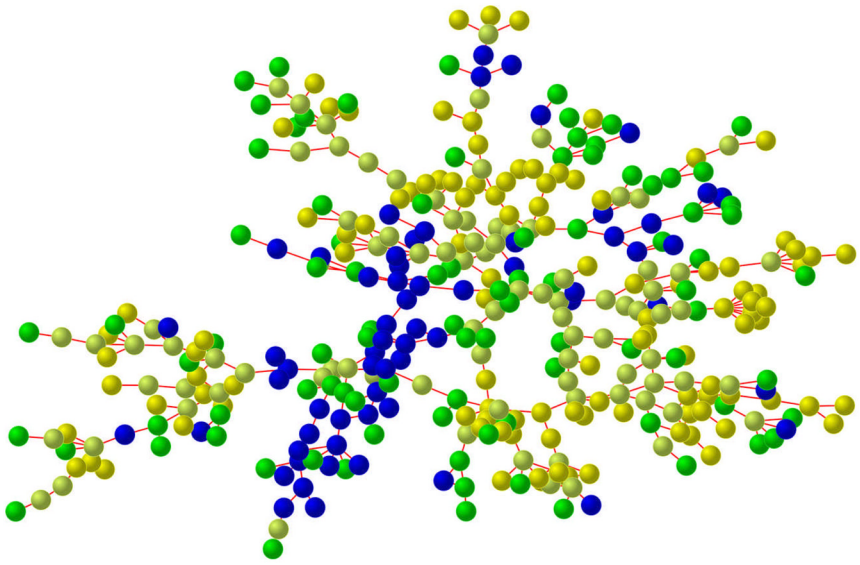

Figure 4. Smiling Clusters in the Online Facebook Network

This graph shows part of a network of friends discerned with the online social network site Facebook in 2007. It contains 353 students; the lines between nodes indicate that the connected individuals were tagged in a photo together. Students who are smiling (and who are immediately surrounded by smiling people in their network) are colored yellow. Students who are frowning (and who are immediately surrounded by such serious looks) are colored blue. Shades of green indicate a mix of smiling and non-smiling friends. It is apparent that blue nodes and the yellow nodes cluster together, indicating large-scale structure of smiling in the online network. Moreover, people who do not smile seem to be located more peripherally in the network.