

# **Minding your Practices: Exposure and Anonymity in Social Network Data**

Caroline Haythornthwaite  
Association of Internet Researchers, 2004, Brighton, UK

## ***Abstract***

This short presentation highlights ethical issues that can result from the collection and display of social network data. The problems include exposure of individuals from whom data has not been collected and thus who could not give permission for their inclusion or exclusion in the data; revelation and exposure of individuals' positions in the network, positions they may not have been aware they held; and exposure of the connectivity among the group as a whole, which may reflect aspects of the group of which they were unaware, and/or that they do not want exposed. The paper draws on some cases of social network data collection by the author to highlight issues that have arisen, and means that have been used to dealing with them.

## **Introduction**

This brief paper introduces some ethical and permission dilemmas associated with the collection and presentation of social network data. This kind of data describes what people are doing with others. Individuals are asked questions such as: Who do you go to for advice? Who do you work with? Who do you socialize with? Who gives you information important to your work? With whom do you discuss important matters? These questions capture the maintenance of relations of various kinds: communication, advice, work, socializing, information, discussion of sensitive issues, or exchange of goods and services. They also capture the direction of these flows: from one person to another, or joint relations like collaboration. Questions can also be asked about the quantity of each relation over a particular timeframe, e.g., How often have you talked to each of member of your team (or class or group) in the last week (month, year)? With such estimates it is then possible to consider only the most frequently maintained relations when looking at the results, showing where strong ties exist among group members.

Social networks are studied in a number of fields: epidemiologists track networks of who contracted a particular disease from whom; sociologists study family trees and kin networks by relations such as who is the child of whom; social psychologists explore the nature of interpersonal relations; health care researchers look at support networks; management researchers look at interconnections among businesses through relations such as who buys from whom, and who sits on the same board together. Major resources on social networks include Wasserman & Faust (1994), Wellman & Berkowitz (1997), Monge & Contractor (2003), Wasserman & Galaskiewicz (1994), and Nohria & Eccles (1992). Social network studies of the Internet and CMC include studies of scholarly communication (Haythornthwaite & Wellman, 1998; Koku, Nazer & Wellman, 2001; Walsh & Bayma, 1996; Walsh et al, 2000), online education environments (Haythornthwaite, 2000, 2001, 2002), help-giving in organizations (Constant, Kiessler & Sproull, 1996), and interconnectivity in websites (Park, 2004) to name a few.

Social network data such as these provide information about an individual's relations with others in their world, and, when collected from all members of a team, groups, etc., reveal information about relations among group members. Ethical considerations about data from each type of data collection are discussed next.

## **Ego networks**

Data can be collected from individuals to describe their *ego network*, i.e., what the world of relations look like from their view to others. In this case, the primary exposure is that of the individual and their relations with others. Others are likely to be named, but in such cases, the researcher is not as interested in specifically who else is involved, but rather the view from the individual, and perhaps the roles others occupy. Thus, the data of interest is things like: How many people are in ego's network? Who – parent, friend, co-worker, boss – is most likely to give money, small services, advice, social support? Who – in terms of role – are they most likely to work with, socialize with, take vacations with, etc.? Pseudonyms can be used as in any other kind of data collection to disguise the identity of the person who responded, and also need to be used for the people they identify.

Sometimes researchers are interested in the connections among members of an ego network. In these cases, ego's are asked how well people they have named know each other, whether those people work together, socialize, go on vacations, etc. This *egocentric* network reveals the interconnectedness of individuals' networks. Some exposure of others begins to show here, but most researchers are not interested in the specifics of such connectivity, but rather in how this plays out across a number of people. Aggregate numbers about egocentric network density are generally what is of interest, and are a good way to maintain anonymity.

However, and here is our first caution, researchers do have in hand data that shows what ego believes to be the interrelation between people who have never been approached for permission to use their name or data, and who can be exposed in pictures of individual's networks. It is here that the social network researcher meets their first research dilemma. They have data that can show information about people who have not given permission for that data to be used.

The normal way to deal with this is first to anonymize the individuals (egos) who have responded. Moreover, it must be recognized that even if the respondent gives permission for his or her name to be used, displaying their name automatically lets others guess about the people they have named. While the general world will not catch the subtleties revealed, those close to ego certainly can. Thus, it is particularly important to maintain anonymity. A second way to deal with this is to report aggregated data only. Even though the researcher could draw a network diagram of each individual's network, it is important to consider whether that is ethical – not with respect to ego, but with respect to the members of ego's network that they have named.

So far, dealing with social network data from individuals does not differ much from ethical data considerations from any kind of data collection. However, the problems become a bit more apparent when we move from data collected from unconnected individuals, to data from connected individuals.

## Whole networks

When data is collected from a whole set of interconnected people, we refer to this as collecting *whole network* data. The idea is to come to understand how the network as a whole is interconnected, what kinds of relations sustain the group, and what kinds of subgroupings are evident within the larger network. One of the perks of whole network analyses is that you can present a visual picture of the network, known as a *sociogram* that shows who is connected to whom. People (or *actors*) appear as nodes in the graph; lines drawn between people show who is connected to whom by the relation of interest.

Sociograms are immediately readable. You do not need to wade through pages of analyses and statistics to discover the interrelations among participants. Sociograms are indeed worth a thousand words. What you see is a clear depiction of the relations among people and how this connects the network as a whole. They are revealing, immediately intuitive, and quite often surprising.

It is the unexpectedness of what is shown that can trouble both individuals, and groups. The data revealed can be disturbing to individuals because it shows how others nominate them in a network: a person who considers themselves the head of the unit may not appear as the sociometric star; people may not be nominated as part of another's important co-workers, when they thought they were; the network may reveal they are isolated from others. But, what equally causes anxiety is that the interrelations will reveal a dysfunctional group, and this will adversely affect superior's opinions of the group. In the cases I am talking about, there were worries that the structures would show something that lead others to cut off the group's research funding. At the time, we were looking at "collaboration," a very vague concept, but one that research groups really want to show they are doing. However, as researchers, we were not as interested in the specifics as in the overall interaction patterns – i.e., we did not have an advance definitions of "successful collaboration." and what it would or should look like on a sociogram.

As a social network researcher, I am aware of what a sociogram may potentially reveal – about individual's positions in the network, and about network relations as a whole. I believe it is my job to be sure that others are aware of this, and are suitably informed to give their consent. This is tricky, because, after all, I do want to collect the data, and I do want to show pictures!

## Getting access

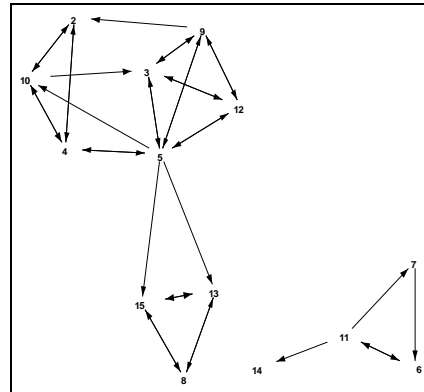
In these studies of collaboration, data were collected from three groups. Although data collection varied somewhat across groups, each included asking about connections among members of the group – collaboration, learning, giving help, etc. For each group, the nature of the work was explained, and this was outlined in permission forms. Each form directed individual's attention to the potential 'dangers' of exposure through the social network diagram, and allowed them to opt out of the study in various ways. Here, for example, is text from one of the study permission forms that directly addresses the network issues:

Although the survey is designed to be anonymous, some presentations of data, notably social network diagrams which graphically represent connections among team members, may make it possible for those close to the team to guess at the identity of individuals (an example of such a network diagram is given below). However, in any reporting of the data your identity will be kept

confidential; neither your name nor the name of your team will be used in reports, publications, etc.

The form also showed this text and diagram:

This is an example of a social network diagram used in publications showing who – among a class of 14 distance learners – worked with whom at least twice a week over the semester.



I chose this particular sociogram to use in my permission form because it comes from data I have collected, and shows some particularly interesting features. One set of actors is disconnected from the others; network stars are apparent in the diagram, as is someone who bridges between two parts of the network (he/she is a ‘cutpoint’, i.e., if removed, the network would fall into separate pieces). I have also used this example elsewhere to illustrate what a network picture can show you (see Haythornthwaite, 2002 in Renninger & Shumar). This diagram itself came from my first studies of whole networks, and revealed to me just how much such a diagram can immediately say about a set of actors. Thus, from past work, I was aware of what people needed to be alerted about, and my considerations of potential respondents has grown from the revelation of this particular diagram.

## Gaining permission

Different groups have reacted differently to the request to collect social network data, and to display it as a graph showing individual data. With one group, the group leader was hesitant to recommend this project to his group members. Without his/her support, it was unlikely that any data could – or even should – be collected. We were eventually successful in getting the consent and support of the group leader, and then from group members, in part by explaining further what we were looking at, and in part by pointing out how long it would be between data collection and funding decisions for the group.

The important point to take away from this group is how much the leader had the welfare of the group in mind, and was making a decision on behalf of the group. Thus, permission was not negotiated with one individual about their exposure, but with the leader about the *group's* exposure. It was not a decision to be made lightly by this group leader – he had his group's interests in mind, not just his own data. Thus, in looking at a whole network, we are really asking for group permission to collect data – data that might reveal structures that might adversely affect the future fortunes of the group.

For a second group, it was necessary for me to not even know about the group's decision or individual's decisions. In this case I used a double blind permission structure, so that even I would not know who had or had not given permission for the use of the group's data. Respondents chose from these two options about what would appear in publications (they were also given the option of not allowing any of the data to be used in publication, but permission was granted for that).

For distribution of aggregated results of analyses in publications in which individuals will not be identifiable

OR

For distribution of results of analyses in publications in which individuals will not be identified by name, but who may be identifiable to others who know the team (e.g., social network diagrams).

They sent their response to a secretary in my department who reported only total permissions to me. The decision of *any one individual* could stop my use of the data in sociograms, and that was the point. If any individual felt uncomfortable, then no data would be shown that way. This is indeed what came out of this permission round. I was *not* given permission to show pictures, but I was allowed to use the data in an aggregate form. Thus, I can look at statistics of the group network, and report on these, but not show the pictures. The point here is that permission needed to be secured individually, and anonymously from every participant. And a system had to be set up so even one individual could change the use of all the data.

The third group was presented much the same information as the others – to the group leader first, and then to each participant. This group had no problem with any of the permission, and exhibited no anxiety about any of the outcomes. In this case, the group leader welcomed the exposure of the activities of a group that had been – in the leader and other's estimation – a roaring success. All members who were interviewed gave permission to use their data even in sociograms. Here the group leader made the decision that it was positive for the study to go ahead, essentially letting individual group members decide about participation, all of whom responded positively. The issue still remains of exposure of people who were not actually interviewed. To date, I have used only aggregate data from this group, and should I find it useful to show a picture, not only will the individuals be anonymized, but also the name of the group itself.

## Summary

There are a number of dilemmas inherent in collecting social network data – some similar to other methods, and some more pressing for network data because of the involvement of non-interviewed network members, and the visual display of the data. The researcher must wrestle with the following dilemmas:

- Understanding what can be revealed from social network data
- Understanding individual exposure in social network diagrams
- Responding to group and individual level anxieties
- Getting group level permission
- Securing individual permission, anonymously if necessary, from all members of the group as necessary

## References

- Constant, D., Kiesler, S.B., & Sproull, L.S. (1996). The kindness of strangers: The usefulness of electronic weak ties for technical advice. *Organization Science*, 7(2), 119-135.
- Haythornthwaite, C. & Wellman, B. (1998). Work, friendship and media use for information exchange in a networked organization. *Journal of the American Society for Information Science*, 49(12), 1101-1114.
- Haythornthwaite, C. (2000). Online personal networks: Size, composition and media use among distance learners. *New Media and Society*, 2(2), 195-226.
- Haythornthwaite, C. (2001). Exploring multiplexity: Social network structures in a computer-supported distance learning class. *The Information Society*, 17(3), 211-226.
- Haythornthwaite, C. (2002). Building social networks via computer networks: Creating and sustaining distributed learning communities. In K.A. Renninger & W. Shumar, *Building Virtual Communities: Learning and Change in Cyberspace* (pp.159-190). Cambridge: Cambridge University Press.
- Koku, E., Nazer, N. & Wellman, B. (2001) Netting scholars: Online and offline, *American Behavioral Scientist*, 44(10),1752-1774.
- Monge, P. R. & Contractor, N. S. (2003). *Theories of Communication Networks*. Oxford, UK: Oxford University Press.
- Nohria, N. & Eccles, R.G. (Eds.) (1992). *Networks and Organizations: Structure Form, and Action*. Boston, MA: Harvard Business School Press.
- Park, H. W. (2004). What is hyperlink network analysis?: New method for the study of social structure on the web. *Connections*.
- Walsh, J. P., Kucker, S., Maloney, N. G., and Gabbay, S. (2000). Connecting minds: Computer-mediated communication and scientific work. *Journal of the American Society for Information Science* 52(14), 1295-305.
- Walsh, J., & Bayma, T. (1996). Computer networks and scientific work. *Social Studies of Science*, 26, 61-703.
- Wasserman, S. & Faust, K. (1994). *Social Network Analysis*. Cambridge, MA: Cambridge University Press.
- Wasserman, S. & Galaskiewicz, J. (Eds.) (1994). *Advances in Social Network Analysis: Research in the Social and Behavioral Sciences*. Thousand Oaks, CA: Sage.
- Wellman, B., & Berkowitz, S.D. (Eds.) (1997). *Social Structures: A Network Approach*. (updated edition) Greenwich, CT: JAI Press.