UWE MATZAT / CHRIS SNIJDERS

The Online Measurement of Ego Centered Online Social Networks

1. Social network analysis and its importance for internet research

Social network analysis (SNA) consists of a set of techniques and theories that analyze the ties between actors and the structure of these ties (WASSERMANN/FAUST 1994). The actors can be individuals, organizations, nations, or any other corporate actors. In principle, the techniques can be and are often used to describe the structure of social relationships between actors. However, they can also be used to describe the structure of ties between websites (e.g. PARK/BARRET/NAM 2002), or the structure of connections between actors in general. The analysis of *social* networks is important because it is known that the structure of social contacts has a fundamental impact on social life and business success. Not only the characteristics of the actor are important, but also the way in which the actor is connected to others and their inter-relations. For example, social networks influence the likelihood of finding a job (GRANOVETTER 1972), career success (BURT 1992), the diffusion of innovations (ROGERS 1995), the success of collaboration between firms (GULATI 1995), opinion leadership (KING/SUMMER 1970), and much more. The often used term *social capital* refers to the insight that an actor's social network can provide (or lack) valuable resources (PORTES 1998) that form a separate factor next to the human capital an actor has.

The internet is not only a tool for the acquisition of information, but it is also useful for the maintenance of existing and the making of new contacts. The massive use of online services, such as FaceBook, Friendster, MySpace, MSN Messenger, and many others that facilitate the enlargement and maintenance of one's social network shows that the internet is very often used for such purposes. Therefore the use of theories and techniques of SNA in the field of internet research promises many valuable insights and is already on its way. For example, some researchers expect(ed) that computer mediated communication and the internet would lead to more equality in participation chances (e.g. BARLOW 1995; DUBROVSKY/KIELSER/SETHNA 1991; GRESHAM 1994). However, Stegbauer and Rausch (1999, 2001), in a study of academic emailing lists, were among the first to show that the matter is more subtle. An analysis of the interrelatedness of the sent email messages revealed a center-periphery structure of different subgroups. The center subgroup had not only a high internal communication density, but was also communicating with many other subgroups. Other members either did not communicate or communicated mostly within the same subgroup (STEGBAUER/RAUSCH 1999). Hence, essentially the same kinds of structures that might arise offline were apparent online. Matzat (2004) showed in a large scale study of the use of emailing lists by university researchers that their utilization increased the informal networks of the users. However, this did not reduce inequality in the contact opportunities between those who had many and those who had only a few contacts. It seems that SNA reveals inequality online more than one had thought. In a business context, Tsang and Zhou (2005) and Lyons and Henderson (2005) show that opinion leadership for products can also be studied in online environments. Finally, Wellman and colleagues argue that the internet would contribute to a *networked individualization* of modern societies (WELLMAN 2000; 2001). By this is meant that the use of the internet would enlarge the individuals' social networks and that in general the networks would become more dispersed and less dense since the different contact persons of the internet user would be rather unlikely to know each other (BOASE/WELLMAN 2006). If this claim turns out to be true then online communication would have a tremendous impact. However, the claim can only be tested by using SNA.

In SNA there is a fundamental distinction between two types of social network data. On the one hand, it is possible to measure the complete network of a well-defined set of actors. A classical example would be the friendship network among pupils of one school class or the communication network of members of an emailing list. On the other hand, one can measure the network from the perspective of one actor (ego). For example, in a general population survey one can ask every respondent about his or her friendship network and perhaps the relationships between these friends. The first type of data is called *complete network data*, the second type is called *ego centered network data*. Both approaches have their strengths and weaknesses. While ego centered network data capture only limited parts of the overall network, this type of data collection can be combined with large scale random sampling techniques. Moreover, the data can be analyzed with the standard techniques of social science data analysis (WASSERMANN/FAUST 1994) which makes the collection of ego centered network data very appealing to social scientists. The problem is that the collection of such data is rather time intensive and requires the focused attention of the respondent, as will be shown in the next section. It is therefore an open question whether ego centered online networks can be measured by means of an online survey. This chapter addresses the question to what extent the online measurement of ego centered social networks is possible, what may go wrong, and under what conditions it is advisable or not advisable to measure them online. Moreover, we illustrate some of the insights that the analysis of ego centered online networks can deliver. In the second section we present the procedure by which ego centered social networks are collected in traditional paper-and-pencil surveys. Furthermore, we briefly summarize the limited knowledge about what may happen to the quality of the data when we try to measure ego centered networks online. The third section presents some background information about a study that conducted online measurements of ego centered online social networks of users of 45 online communities. In the fourth section we show some insights based on the empirical analysis of ego centered social networks. We finish the chapter by drawing some conclusions about what the critical issues of the online measurement of online social networks are and under what conditions it is (not) advisable to collect the data online. At the same time, we emphasize that while online measurement of social network data is already on its way (see MARIN 2004), the knowledge about what the transition from offline to online measurement means to the data quality is still limited and under continuous revision.

2. Measuring ego centered social networks

Traditionally, the measurement of ego centered social networks is done with the help of an interviewer who is available for assistance and who can motivate the respondent to continue with the answering procedure. We first present the traditional data collection method to show why the measurement is more difficult than the measurement of demographic or other characteristics. Thereafter, we present the limited knowledge about what happens with the data quality when we transfer the data collection procedure to an online survey.

2.1 *The standard measurement procedure*

The by far most often used method to collect ego centered network data was proposed by Burt (1984). It is regularly used in the US General Social Survey since 1984 (see e.g, MCPERSHON/SMITH-LOVIN/BRASHEARS 2006). The data collection method proceeds in three steps. In the first step the respondent (ego) is confronted with a so-called *name generator* that asks him to list a limited number of individuals (*alteri*) with whom he is in a well-defined, usually close relationship. Such a relationship typically includes, for example, discussing work decisions, visiting regularly, discussing personal problems etc. For instance, for measuring an individual's personal discussion network, Burt (1984) proposes the following name generator. »*From time to time, most people discuss important personal matters with other people. Looking back over the last six months, who are the people with whom you discussed an important personal matter? Please just tell me their first names or initials.*« When the respondents have listed the initials of a number of persons, these persons (alteri) are regarded as being members of the respondent's social network. The interviewer usually writes down the initials of a limited number of alteri. For large scale general population surveys Burt (1984) proposed limiting the list to 5 alteri. The results of earlier social network studies suggest that respondents usually mention 0-8 alteri with an average of 3 in reaction to questions about close relationships. In the second step a number of questions about the characteristics of the cited alteri and about the relationship of the respondent with every alter of his network are asked. These are called *name interpreters*. For example, for every alter there are questions about the age, gender, educational background, and the strength of the relationship between ego and alter. In the third step data about the relationships between the different alteri within ego's social network are collected using a so-called *inter-alter response matrix*. The respondent is asked to describe the relationship between a specific pair of alteri. For example »*Think about the relationship between* (COLUMN NAME) *and* (ROW NAME). *Would you say that they are strangers, just friends, or especially close?*« The question is then repeated for each pair of alteri named by the respondent. In a shorter version of this procedure respondents are simply asked to list those pairs who know each other (BURT 1984). That shortcuts to this procedure are useful can be easily understood by considering a respondent who has given ten alteri in the name generator question. If there are, say, four questions about each alter in the name interpreters, we get a total of 40 (=10 times 4) items to be answered. The response matrix asks for another 45 items (=10*9/2). Taken together with the initial name generator, this implies answering 95 items (=10+40+45). And that only gives the researcher the ego-network and nothing else. Such network data are used to calculate indices that describe characteristics of the respondent's social network, such as its size, density, or the level of homophily in the network. In network analyses the network characteristics are then used to correlate with various other measurements, such as career success, or they are to be predicted by other characteristics of the respondent, such as age, gender etc.

The outcomes of the measurements when carried out using a paper-and-pencil-with-interviewer context are known to be sensitive to details of the measurement procedure. We know that the measurements do not provide a perfect picture of the respondent's recent interaction (BERNARD/KILLWORTH/SAILER 1982). The literature demonstrates that respondents are not good in recalling specific interactions or interactions that took place within a specific time boundary. However, respondents are reasonable good at reporting their typical, stable social relations (MARSDEN 1990). There is some bias and error in the respondents' recall of their relations, but we have some information about the types of biases that emerge. There are systematic differences between reported and observed relationships of respondents. For example, when confronted with a *name generator*, it is likely that a respondent mentions his or her frequent and close contacts, contacts that are more central in the network, and multiplex relationships rather than his or her infrequent, distant, less central or simple instrumental contacts (KOGOVSEK/FERLIGOJ 2004; MARIN 2004). There exist different name generators for different types of social networks. Of course, different name generators can lead to different network sizes (MARSDEN 1990; LOZAR/MANFREDA/

VEHOVAR/HLEBEC 2004), but they all tend to focus the respondent's attention to the close, intimate ties (LOZAR MANFREDA/VEHOVAR/HLEBEC 2004). Also, there is a high test-retest stability of the names reported in the name generators. For example, for Burt's (1984) »discussion of personal problems« name generator, there was an overlap of about 70% of the mentioned names within an interval of 4 weeks. Since such discussion networks are not perfectly stable, this is regarded as relatively high (e.g., MARSDEN 1990).

The quality of the data obtained by the *name interpreter,* measured by the degree of overlap between the reports of ego and alter on alter's characteristics, tends to be high for socio-demographic characteristics of the alteri, but much lower for attitudes or opinions (MARSDEN 1990). The quality of the data on the characteristics of the relationships between ego and alter, measured by the degree of concordance in the reports of alter and ego, tends to be particularly high for close ties and general types of interaction. This is known for characteristics of the relationship such as the frequency of interaction, its duration, and its intensity (MARSDEN 1990). We do not know of any study that has tried to assess the quality of the data on characteristics of the relationships between the alteri as collected through the *inter-alter response matrix*.

Only limited knowledge is available about the quality of the indices that characterize the network of the respondent as a whole. These network measurements are based on the information obtained by the name generator, the name interpreter, and the response matrix. Marin (2004) compared such network measurements with measurements based on an available list of all alteri. She reports that there is a systematic bias in measurements of network size. High values of density, mean duration, and mean closeness of the network lead to better estimations of network size, whereas for example low density networks tend to be underestimated with respect to network size. While there may be a bias, nevertheless the test-retest correlation of network sizes measured with the help of name generators tends to be high. For example, the 1-week test-retest correlation of friendship network sizes in a study of Fischer et al. (1986) was 0.91, and the 2-day test-retest correlation of the size of social support networks in a study of Barrera (1980) was 0.88. No comparable information is available about the quality of the measurement of network density (see MARSDEN 1990), which is also very often used in studies of social networks.

2.2 *The transition from offline measurement to online measurement*

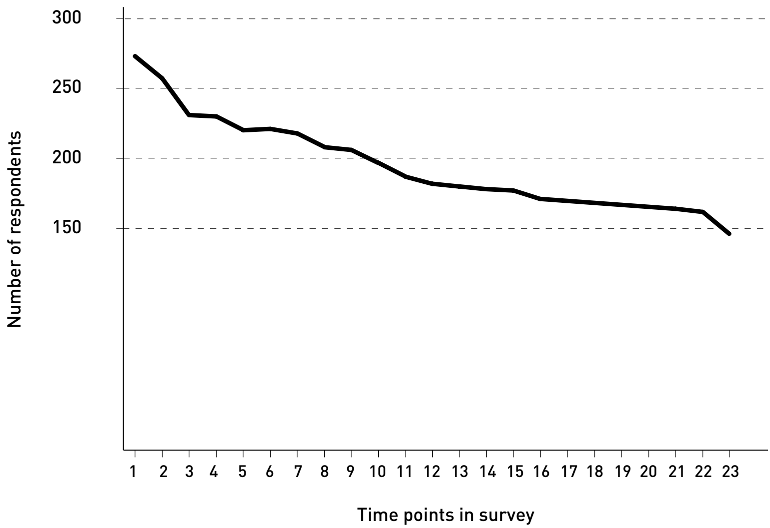
As has become clear by now, the measurement of the respondents' network characteristics is time consuming and demanding for the respondent. Therefore, until recently, almost all network studies were conducted by means of a personal interview. While a web based survey may reduce interviewer effects (LOZAR MANFREDA/VEHOVAR/HLEBEC 2004), it is unclear which other, perhaps disadvantageous effects the lack of social control through the interviewer may have. The presence of an interviewer may be a stimulus for the respondent to continue with the answering procedure. This may contribute to lower proportions of missing values and lower drop out rates during the survey. Hence, the online measurement may lead to a lower data quality with respect to missing values and selectivity. Kogovsek et al. (2002) showed that the collection of ego centered network data is possible by means of a telephone interview. However, while the collection of ego centered network data by means of a web survey is already on its way (see MARIN 2004), we still have only limited knowledge about how this affects the quality of the measured network data.

One of the few studies in this respect is an experimental study by Lozar Manfreda, Vehovar, and Hlebec (2004), who show that the measured network size is dependent on the details of the used name generator. The more placeholders are presented for the recall of relevant alteri, the more alteri are recalled (and consequently the larger the size of the network). Moreover, the larger the network size, the higher the likelihood that the respondent drops outs out of the survey during the second and third part of the network data collection. Matzat and Snijders (2007) used an experimental study to compare the quality of network data obtained by means of a personal interview with the quality of network data obtained by means of a web survey. In the experiment university researchers were asked about their contacts that were useful for the initiation of collaborations with commercial business companies. The authors conclude that respondents in the web survey, when compared to respondents who were interviewed, were more likely to fill out no names at all and more likely to drop out during the filling out of the inter-alter response matrix. If respondents did fill out names, they filled out fewer names during the web survey than during the interviews. There were no large differences with respect to the proportion of missing values. However, the network density measured by means of the web survey was much higher. This arose because respondents during the web survey were much more likely to report that *all* of their alteri were in contact with each other. Matzat and Snijders (2007) interpret the results as being supportive for the hypothesis that during web surveys respondents tend to fill out the response matrix mechanically to save time. They conclude that online measurements of ego centered network data should be conducted very carefully. Given the current state of the art it would be wise to collect social network data online with highly motivated respondents so that the tendency to fill out the questions mechanically can be avoided as much as possible. Furthermore, the online measurement of social networks would profit from knowledge about graphical design elements that provide stimuli to prevent drop out and mechanically answering of less motivated respondents (ibid.).

3. The online measurement of ego centered online social networks: Study design

In the following we present the results of a study that examined to what extent the online measurement of online social networks is feasible and what insights it may provide. Given our previous results, we decided to measure online social networks in what we thought would be relatively active online communities with involved members. In June 2006 we asked a number of members of medical and social support communities on the internet to participate in a short survey concerning their experiences in their community. The questionnaire included questions about the members' use of the community, the relevance of the community in their life online and offline, their satisfaction with the online communication, how they felt about their community, and many other aspects. All in all, 273 members of 45 communities participated in the survey, of which 146 (53%) filled out the questionnaire completely. We approached only English language communities that had at least 10 messages posted within the past 4 weeks. Filling out the questionnaire took about 25 minutes. Name generators (part 1 of the network measurement procedure), name interpreters (part 2), and the response matrix questions (part 3 of the network measurement procedure) were placed at the end of the questionnaire. This made sure that respondents who answered questions about their online network were rather motivated – it is well known that respondents who have progressed relatively far in the survey tend to finish it. By this we hoped to reduce the respondents' tendency to save time by filling out less names, by dropping out during the network measurements, or by answering the questions mechanically. Figure 1 shows the number of respondents still in the questionnaire, per time-point. Roughly, each time point represents a page, except for the last two: between time-points 22 and 23 are *all* network questions.

FIGURE 1
The number of non-dropout respondents per time-point



As can be seen in Figure 1 (and as is a standard finding in online surveys), there is a relatively large number of drop-outs in the beginning. For the network questions (between time-points 22 and 23) we indeed see that there is a serious increase in the number of drop-outs of a size that is roughly similar to the drop-out rate in the first couple of pages. From those still in the survey before the network questions 90% (146 from 162) complete the network questions. Hence, 59% of the respondents (162 out of 273) complete the questionnaire up to the network questions. The network questions themselves reduce the response rate to 53% (146 out of 273). Note that, as we mentioned above, *all* social network questions occur between time-points 22 and 23. So we actually measure the drop-out over several separate pages between these time-points: two pages of name generators (one for friends and one for acquaintances), two pages of name interpreters, and the inter-alter response matrix. All in all, this shows that when put at the end of a survey, drop-out rates *per page* are roughly in line with the general drop-out rate per page across the survey as a whole.

We did not include any fancy graphical elements in the design of the name generators, name interpreters, or the response matrix. Therefore the graphical layout of the network part of the questionnaire very much resembled the traditional layout of a paper-and-pencil questionnaire that is used for data collection supported by an interviewer. Figure 2 shows what the name generators looked like.

FIGURE 2

The name generator

Please note the names, unique initials, or nicknames of **at most 5 friends** you made through **your community** in the fields below. You can use pseudonyms as long as you can remember who the real person behind the pseudonym is. Please mention every name only once.

1	
2	
3	
4	
5	

Please note the names, unique initials, or nicknames of **at most 5 acquaintances** you made through **your community,** that you do not consider to be personal friends, in the fields below.

1	
2	
3	
4	
5	

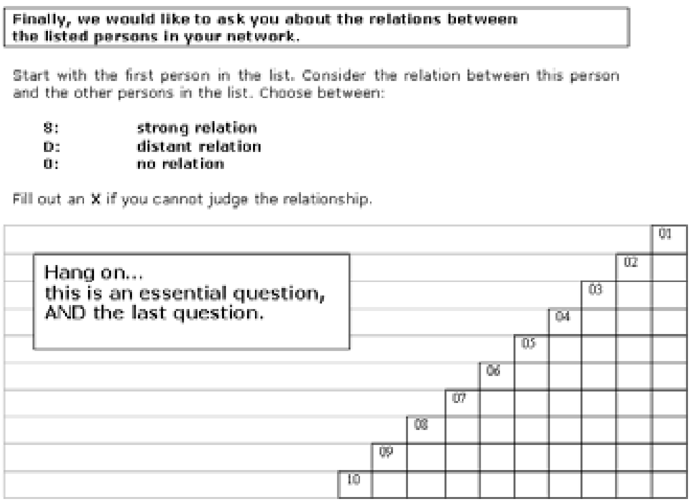
Note: the two name generators were presented to the respondent at separate web pages

The used name generators gave the respondent the opportunity to fill out up to 10 names: the pseudonyms of up to 5 friends and 5 acquaintances (s)he has made through his or her online community. We explicitly asked the respondent not to include any romantic relationships. How long respondents spend on the social network questions is of course dependent on how many contacts they listed. On average the network questions took about one and a half minute per tie. The average time used for the social network questions was 4:22, with 75% of the respondents between 0:25 and 5:55 (median=2:43; 95% between 0:12 and 14:23).

The name interpreter questions consisted simply of questions about each contact and are not shown here. In Figure 3 we do show what is usually considered the culprit in the measurement of social networks: the inter-alter response matrix.

FIGURE 3

The inter-alter response matrix


Finally, we would like to ask you about the relations between the listed persons in your network.

Start with the first person in the list. Consider the relation between this person and the other persons in the list. Choose between:

S: strong relation
D: distant relation
0: no relation

Fill out an X if you cannot judge the relationship.

Hang on...
this is an essential question,
AND the last question.

01 02 03 04 05 06 07 08 09 10

Note: The response matrix was presented on the last page of the web survey. On every row of the matrix the respondent would see the pseudonym that he had filled out earlier

4. The online measurement of ego centered online social networks: results

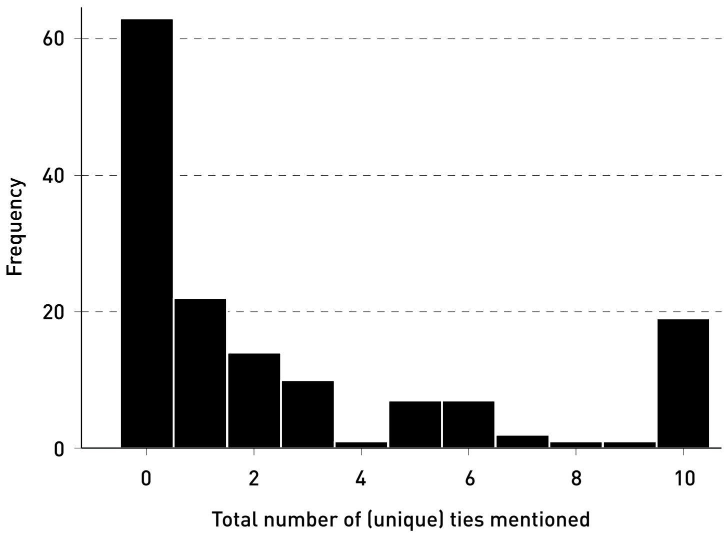
The fact that the study does not apply a random sampling procedure and the fact that we typically have the more involved members answering the network questions makes it hazardous to draw conclusions about the ›average‹ member of the selected online communities. Nevertheless, we can analyze what the characteristics of the online networks of the highly motivated community members look like, whether there are differences in the online networks between members of different communities and with what advantages and disadvantages specific networks characteristics are associated. We first describe some general characteristics of the data.

The mean age of the respondents is 39 years, 75% of the respondents are between 29 and 49 years old (median=40 years, 95% between 18 and 58). From the 160 respondents who indicated their marital status 54% are married, 32% are single, 13% have a girl- or boyfriend, and a little bit more than 1% are widowed. Out of the 164 who informed us about their country of origin 51% come from the US, 20% from India, 9% from the UK, 4% from Canada, 2% from Australia, 2% from New Zealand, and the others come from many different countries all over the world. Within the group of 160 respondents who gave information about their educational background the mean number of years of education is 17, and 75% have at least 14 years of education (median=17, 95% between 7 and 21 years). There are 101 female respondents, 59 are male, and 113 did not indicate their gender. A large majority (76%) of the respondents has been using the internet for at least 4 years and 35% for at least 10 years. The average community member in our sample spends daily about 2-3 hours on the internet, and has been a member of his or her community between 1 and 2 years. One in three has been a member for more than 2 years. On average, the respondent visited his or her community during the past 3 months between 4 and 7 times a week. A minority of 28% visited it more than once a day. In our sample, the wish to receive information and the intention to find social support are both prominent motivators for becoming a member. On the question ›what are your reasons for membership in the community?‹ 66% of the respondents mentioned ›searching for information‹ and 52% mentioned ›looking for support‹. Many respondents attach a value to the relationships with other members. Only 8% indicated that it is not that important for them to have good relationships with other members, 17% explicitly denoted they are indifferent and about 75% agreed that they care to some extent. The wish to build up or to maintain good relationships with other members of the community is being realized for many respondents. To the statement that they have the feeling to be able to make new contacts in the community, 60% agreed. Finally, about

55% of the 159 respondents who answered the question whether they had found new contacts (acquaintances or friends) in the community said that they had done so.

We now consider the different network measures in some detail. We show that there is considerable variability between individuals across the network measures and that the network measures vary in a meaningful way with other characteristics of the members. Let us first look at how many ties the members mention in the two name generators (friends and acquaintances). Figure 4 shows the distribution.

FIGURE 4
Distribution of number of ties mentioned (n=146)



We see that most members either mention zero (43%), one (15%), or ten (13%) contacts that they made through their online community. These percentages change somewhat but not much if we would only consider members who have joined the community more than 3 months ago. When we look at which persons are more likely to have ties at all, we find several correlations that are interesting.[1] First, we find that members who

1 For all results mentioned here, we use (versions of) multivariate regression analyses. Tables with the detailed results are available on request from the authors.

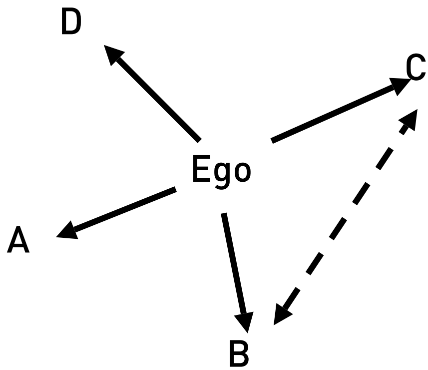
tend to agree with the statement that their community is ›a social group‹ are more likely to have social ties (regarding the community as a social group measured on a 6-point scale; an increase of one standard deviation on the scale leads to an increase in the probability to have ties of 24 percentage points). With a social group we mean that the community is not only used as a source of information but adds to the social well-being of the member as well.[2] We also find that members who are more active are more likely to have social ties (activity is measured as a scale of five different items; an increase of one standard deviation on the scale leads to an increase in the probability to have ties of 16 percentage points). Finally, we find that members who are from the US are less likely to have ties to other members of their community, a difference of 30 percentage points. No differences are found between men and women, or between age groups. When we run a similar analysis trying to explain which members are more likely to have 10 (or more) contacts, we find again that the active members are more likely to have this, just as those who consider their community a social group. No difference for US members is now found. In an analysis on the (log of) the total number of ties we also see similar results. Active members in online communities that they consider a social group are the ones who have more ties. These results make sense intuitively and can serve as an indicator that the name generators have external validity: they seem to measure what they are supposed to measure.

Next, we consider the density of the ego-networks. To explain what is meant by density in this case in more detail, consider Figure 5.

As can be seen in Figure 5, Ego has mentioned 4 alteri, A through D, and has mentioned in the inter-alter response matrix that B and C know each other well, and the rest do not know each other at all. We define the density of the network as the number of ties that are present between the alteri, as a percentage of the total number of ties that could be present. In Figure 5 only one tie exists (between B and C), whereas there could have been six ties (AB, AC, AD, BC, BD, and CD), so the density of the alter network equals 1/6. In Figure 5 we show binary relations (the relation exists or not), but in our data we can even value these networks. As it turns out, in our data calculating density using the valued or the binary data does not matter much.

2 We measure this as a scale score of six 6-point items asking whether respondents feel they are ›well known‹, ›appreciated‹, ›are challenged‹, ›have fun‹, ›can make contacts‹, and ›can maintain satisfying relationships‹.

FIGURE 5
An example ego-network



The correlation between the two measures is 0.92. We use the valued density measure in the findings that we report here. The first thing to note is that we have fewer data points. All respondents who mentioned zero ties have missing values on the density measure by definition. Taken together, we have density data for 62 respondents, which obviously makes finding meaningful correlations harder. The mean density is 0.17, and 75% of the density values lies between zero and 0.25 (median 0.18, 95% between 0 and 0.42). Note that these are relatively high values. The average density among the respondents' online network is thus somewhat higher than the density of the example network in Figure 5. Apparently, there are community members out there who not only have ties to others but whose contacts themselves are connected to each other to a reasonable degree. Hence, the general image of individuals being connected online to other unconnected individuals (cf. BOASE/WELLMAN 2006) is refuted in our data.

Because the potential number of ties between alteri increases fast with the number of alteri, it is likely to assume that the actual density decreases when more alteri are mentioned by the respondent. Actually the bivariate relationship is curvi-linear: the average density increases from 0.13 to 0.35 when the number of alteri increases from 2 to 7, after which it decreases to 0.18 for those with 10 ties. An analysis on the 62 cases from which we have density data only shows some support for the fact that US-citizens tend to have denser networks (+0.07 increase in density; p=0.03) and the active members also have denser networks (+0.04

increase in density for an increase of one standard deviation in the activity measure; p=0.08). Another effect that is marginally significant but worth the mention is that the density is higher on community sites that have many ›Messenger-like‹ characteristics, such as chat boxes, the possibility of avatars, the use of color, and the possibility to create online ›buddies‹. The effect is only at the p=0.10 level, but nevertheless interesting because when it can be substantiated by further research, it has important implications for those maintaining a site. When we tentatively calculate the average density per community and treat this as a community characteristic – an admittedly crude way to increase the number of data points over which we can analyze – we indeed find that being a member of a community with many possibilities to personalize interaction increases the density of one's online network somewhat. This gives some support for the idea that there are systematic differences between the online communities with respect to the density of the online networks they create.

5. Summary and conclusion

In this chapter we examined to what extent it is possible to measure ego centered online social networks by means of an online survey and its implications for the quality of the network data. The measurement of social networks is time-consuming and demanding for the respondent. Therefore, traditionally this is done by means of a paper-and-pencil survey and with the help of an interviewer who can explain the procedure, motivate the respondent, and exert some implicit social control. This prevents respondents from making mistakes and ensures that respondents with large networks (who necessarily need more time to fill out the questions) do not drop out of the survey. The standard data collection procedure leads to some known biases in the measurement of social networks. Nevertheless, respondents are known to be able to recall their typical stable and more intimate relationships in a reasonable way. Much less is known as to what happens with the quality of the measurements when we try to collect ego centered network data with the help of a web survey. The limited existing evidence suggests that during a web survey respondents try to save time by mentioning fewer or no social contacts when compared to respondents who answer the questions during a personalized interview. Moreover, there is a tendency for average respondents to fill out the response matrix that measures the ties between the contacts of the respondent rather mechanically. This makes the online measurement of ego centered social networks even more difficult than it already is. For this reason, online collection of social network data can best be done with highly motivated respondents who are less inclined to quit the survey before they reach the end and who tend to invest more time in the completion of the questionnaire. We presented the results of a study that illustrates how the data collection can be conducted in a way that resembles the traditional paper-and-pencil questionnaire. The respondents were members of 45 different social support online communities and we asked them questions about their friends and acquaintances they made within their online community as well as about the relations between these friends. Since the network questions were placed at the end of the questionnaire it was likely that those respondents who actually started answering them are likely to be more motivated because they invested already about 15 minutes in filling out the survey questions.

The results show that among these highly motivated respondents the drop out rate per question is roughly comparable to the drop out rate during the other non-network parts of the questionnaire. Moreover, the found network characteristics seem to have some face validity. Those online community members who are more active and who tend to regard the community not only as a source of information, but as a ›social group‹ have made more contacts. Moreover, we find that the online networks are not as sparse as one might expect based on the literature (cf. BOASE/WELLMAN 2006). Again, more active members tend to have more dense online social networks and there is a curvilinear relationship between the size and the density of the online network, as one might expect. Finally, there seem to be systematic differences between the online communities with respect to the density of their members' online networks. In some online communities members have a higher likelihood of developing a somewhat denser online network. All in all, the results increase our confidence in the possibility to measure online social networks through web surveys for highly motivated respondents. Additional research should show precisely what the best way to gather such data is. For sure the increased graphical possibilities in online surveys might be a way forward: presenting the different questions in a smoother, less tedious or clever shorthand way would certainly do much to improve the reliability of the answers and completion rates.

Matters are more complicated for less motivated respondents. At the moment it is unclear whether collecting ego centered social network data of less motivated respondents by means of a web survey is feasible, at least in such a way that the quality of the data does not seriously suffer. Here we urgently need more knowledge, about design elements of a web survey, but also about motivating research strategies, such as approaching less motivated respondents in a special way to make sure that they do not loose their interest too early. Such knowledge would certainly be valuable and could facilitate online survey strategies while incorporating questions about the social structure of relationships, both online and offline.

References

BARLOW, J. P.: Property and speech: Who owns what you say in cyberspace? In: *Communications of the ACM, 38,* 1995, S. 19-22

BARRERA, M.: A method for the assessment of social support networks in community survey research. In: *Connections, 3,* 1980, S. 8-13

BERNARD, H. R.; P. D. KILLWORTH; L. SAILER: Informant accuracy in social network data V: An experimental attempt to predict actual communication from recall data. In: *Social Science Research, 11,* 1982, S. 30-66

BOASE, J.; B. WELLMAN: Personal relationships: On and off the internet. In: VANGELISTI, A.; D. PERLMAN (Eds.): *Handbook of Personal Relationships.* Cambridge 2006, pp. 709-724

BURT, R. S.: Network items and the General Social Survey. In: *Social Networks, 6,* 1984, S. 293-339

BURT, R. S.: *Structural Holes*. Cambridge, MA [Harvard University Press] 1992

DUBROVSKY, V. J.; S. KIESLER; B. N. SETHNA: The Equalization Phenomenon: Status Effects in Computer-Mediated and Face-to-Face Decision-Making Groups. In: *Human-Computer Interaction, 6,* 1991, S. 119-146

FISCHER, J. L.; D. L. SOLLIE; K. B. MORROW: Social networks in male and female adolescents. In: *Journal of Adolescent Research, 6,* 1986, S. 1-14

GRANOVETTER, M.: The strength of weak ties. In: *American Journal of Sociology, 78,* 1972, S. 1360-1380

GRESHAM, J. L. JR.: From Invisible College to Cyberspace College: Computer Conferencing and the Transformation of Informal Scholarly Communication Networks. In: *Interpersonal Computing and Technology: An*

Electronic Journal for the 21st Century, 1994, S. 37-52

GULATI, R.: Does Familiarity Breed Trust – the Implications of Repeated Ties for Contractual Choice in Alliances. In: *Academy of Management Journal, 38,* 1995, S. 85-112

KING, C. W.; J. O. SUMMERS: Overlap of Opinion Leadership Across Consumer Product Categories. In: *Journal of Marketing Research, 70,* 2006, S. 43-50

KOGOVSEK, T.; A. FERLIGOJ: The quality of measurement of personal support subnetworks. In: *Quality and Quantity, 38,* 2004, S. 517-532

KOGOVSEK, T.; A. FERLIGOJ; G. COENDERS; W. E. SARIS (2002). Estimating the reliability and validity of personal support measures: Full information ML estimation with planned incomplete data. In: *Social Networks, 24,* 2002, S. 1-20

LOZAR MANFREDA, K.; V. VEHOVAR; V. HLEBEC: Collecting ego-centred network data via the web. In: *Metodolzski zvezki, 1,* 2004, S. 295-321

LYONS, B.; K. HENDERSON: Opinion leadership in a computer-mediated environment. In: *Journal of Consumer Behavior, 4,* 2005, S. 319-329

MARIN, A.: Are respondents more likely to list alters with certain characteristics? Implications for name generator data. In: *Social Networks, 26, 2004, S.* 289-307

MARSDEN, P. V.: Network data and measurement. In: *Annual Revue of Sociology, 16,* 1990, S. 435-463

MATZAT, U.: Academic Communication and Internet Discussion Groups: Transfer of Information or Creation of Social Contacts? In: *Social Networks,* 26, 2004, S. 221-255

MATZAT, U.; C. SNIJDERS: *Does the Collection of Ego-Centered Network Data on the Web reduce the Data Quality? An Experimental Comparison of online versus offline Data Collection*. Working Paper. Eindhoven University of Technology. 2007

MCPHERSON, M.; L. SMITH-LOVIN; M. E. BRASHEARS: Social isolation in America: Changes in core discussion networks over two decades. In: *American Sociological Review, 71,* 2006, S. 353-375

PARK, H. W.; G. A. BARNETT; I.-Y. NAM: Hyperlink-affiliation network structure of top web sites: Examining affiliates with hyperlink in Korea. In: *Journal of the American Society for Information Science and Technology, 53,* 2002, S. 592-601

PORTES, A.: Social Capital: Its Origins and Applications in modern Sociology. In: *Annual Review of Sociology, 24,* 1998, S. 1-34

ROGERS, E. M.: *Diffusion of Innovations*. New York/London/Toronto/Sydney/Tokyo/Singapore [The Free Press] 1995

STEGBAUER, C.; A. RAUSCH: Ungleichheit in virtuellen Gemeinschaften. In: *Soziale Welt, 50,* 1999, S. 93-110

STEGBAUER, C.; A. RAUSCH: Die schweigende Mehrheit – ›Lurker‹ in internetbasierten Diskussionsforen. In: *Zeitschrift fuer Soziologie, 30,* 2001, S. 48-64

TSANG, A. S. L.; N. ZHOU: Newsgroup participants as opinion leaders and seekers in online and offline communication environments. In: *Journal of Business Research, 58,* 2005, S. 1193

WASSERMAN, S.; K. FAUST: *Social Network Analysis*. Newbury Park CA [Sage] 1994

WELLMAN, B.: Changing Connectivity: A Future History of Y2.03K. *Sociological Research Online, 4*. URL (retrieved 22 December 2006): http://www.socresonline.org.uk/4/4/wellman.html (2000)

WELLMAN, B.: Physical Place and Cyber Place: The Rise of Personalized Networking. In: *International Journal of Urban and Regional Research, 25,* 2001, S. 227-252