ELSEVIER

Social Networks 19 (1997) 303–323

SOCIAL
NETWORKS

Eliciting representative samples of personal networks

C. McCarty, H.R. Bernard, P.D. Killworth, G.A. Shelley, E.C. Johnsen

University of Florida, Florida, USA

Abstract

In this paper we introduce and evaluate a method for eliciting a representative sample of total personal networks. First names were used as a cue to elicit a sample of 14 alters from 712 respondents through a telephone interview. Network characteristics for each respondent were calculated as averages and proportions across the 14 alters. These were compared to other studies using more specialized network generators. Our method produced results which are logically consistent with those expected from a generator that elicits a sample from the total rather than a specialized subset of the total network. The proportions of kin relations, average tie strength and frequency of contacts are found to be lower than network generators designed to elicit networks of social support. Given our conclusion that the sample is representative of the total network, we examine the varying characteristics of respondents and their networks based on the domination of a particular relation type in their network. This analysis provides answers to such questions as 'What characteristics of respondents account for the proportion of family relations in their network?' and 'What are the similarities between respondents whose networks are made up of mostly work-related relations?' © 1997 Elsevier Science B.V.

1. Introduction

Most studies of networks begin with a bounded social group, such as an organization. The connections among all pairs of people in the group are then mapped by asking respondents to indicate the existence of (or to rank or rate) some social or emotional tie ('owe money to' or 'like to work with,' for example). Studies of such *complete networks* produce matrices, one matrix for each connection mapped. The matrices are then examined for structural features such as density and centrality. These quantities, in turn, can be related to outcome variables like profit, productivity, affect, etc. Studies of complete networks have yielded information about communication patterns in organizations and about the formation and decay of structural subgroups.

PII S0378-8733(96)00302-4

A smaller set of network studies begin with an individual, whose connections are mapped to others in his or her social universe. Studies of such *personal networks* are almost always about subsets of network alters rather than about total networks. These studies rely on questions like 'who do you go shopping with?' and 'who do you talk to about important matters?' and have yielded information on how people tap human resources for everything from finding an abortionist (Lee, 1969), coping with personal disaster (Williams, 1995), to getting a job (Granovetter, 1974).

Few studies have been done on *total* personal networks — and with good reason. Depending on the definition of a tie, total personal networks run from about 250 alters (Killworth et al., 1984) to about 5000 (Pool and Kochen, 1978, Freeman and Thompson, 1989, Killworth et al., 1990). Even at the low end of this range, it is impossible to collect data on each alter in the personal networks of respondents. Indeed, studies of total personal networks have so far been restricted to estimating network size (Pool and Kochen, 1978, Bernard et al., 1989, Killworth et al., 1990, Freeman and Thompson, 1989, Johnsen et al., 1995, Killworth et al., 1990, Killworth et al., in press) and to describing network components (McCarty, 1992).

There are, however, at least three reasons for studying the characteristics of total personal networks.

1. They will add information to studies of network subsets. Wellman (1979), for example, found that 50% of the people we 'feel closest to' are kin. Finding, as McCarty (1992) did, that kin comprise 27% of respondents' total list of network alters adds information to Wellman's finding about the list of close alters.
2. They will help us assess the possible impact of network features on variables like wealth accumulation and longevity. By network features we mean things like density, centrality, and so on, as distinct from features of network alters (like average age, average income, and so on). Representative samples of respondents' networks would support many analyses that are now not possible.
3. They will help us estimate the average size of personal networks. Cross-cultural comparisons of networks will be more useful when we can estimate accurately the size of networks. Estimating the size of uncountable populations (like the number of people who are HIV-positive, or the number of rape victims in a city) will also be more accurate (Bernard et al., 1989, Laumann et al., 1993, Killworth et al., in press).

In this paper we introduce and evaluate a method for eliciting a representative sample of total personal networks.

2. Methods

It is reasonable to assume that no two people have exactly the same list of network alters. Thus, to get a representative sample of any respondent's network, we need a cue that stimulates more-or-less unbiased recall of alters in that network.

We have tested the possibility that first names provide such a cue. We selected a list of 50 first names from the 1993 University of Florida registrar's records, selecting names that were common among both Black and White students. For example, few

White women are named Letisha, and few Black women are named Megan. We also tried to select names that were popular among different age groups. For example, many college-aged women are named Jennifer, while almost none are named Helen (the reverse of the situation 50 years ago). We chose the registrar data because it allowed us to see how names varied by age and race so that we could pick names which were relatively free of bias. A similarly structured data set from the US Census 5 million person sample would be a better choice, but was unavailable at the time of data collection.

From December 1993 to March 1994, we interviewed by telephone 1 525 respondents selected by random digit dialling from the population of all Florida households with a telephone. Of those contacted, 52% (1 525) agreed to do the survey. Respondents were randomly selected from all adults in the household using the next-birthday method. With this method, the person who answers the phone is asked to get the adult who lives in the household who has the next birthday. If that person is not available, and arrangements to call them back cannot be made, the household is excluded from the sample.

Respondents were told that they would hear a list of first names. (The list was comprised of alternating male and female first names.) Respondents were asked to tell us when they heard a name that matched someone they knew. Interviewers then read the list of 50 names to respondents, and when respondents said they recognized a name, the interviewers asked some questions about the network alter and the respondent's relationship to the alter. For this survey 'knowing' meant that the alter would recognize the respondent by sight or by name, that the respondent could contact the alter and that they had contact within the past two years.

Interviewers continued reading the list of first names until respondents recognized 14 names or until the list of 50 names was exhausted. In all, 712 respondents (47%) generated 14 alters.

For each of the 14 alters, we asked respondents the *age* of the alter; *how they knew* the alter (coded by interviewers as one of 23 categories); *how well they knew* the alter (on a scale of 1 to 5); *how much they knew about* the alter (on a scale of 1 to 5); *how long they knew* the alter (a 5-point ordinal scale from 'less than a month' to 'more than 10 years'); and *how often they had contact with* the alter (a 5-point ordinal scale from 'every day' to 'less than once a year').

For analysis, these characteristics were averaged across the 14 alters for each respondent. Thus, respondents were assigned an 'average age of their alters,' an 'average amount of time they knew their alters,' and so on. This part of the interview, including selection of 14 alters and asking questions about each alter took 15 min, on average, per respondent.

In the next section of the interview, 20 of the 91 possible pairs of alters $[(14 \times 13) \div 2]$ were presented to the respondent who indicated (yes or no) whether the members of the pair knew one another. We paired particular alters by sequence to simulate a random pairing. For example, we always asked if the first alter chosen knew the eleventh alter chosen. Given varying patterns of name recognition these were different names for different respondents. Density measures were estimated by dividing the number of reported ties by 20. Most respondents completed this part of the interview in less than 2 min.

3. Respondent characteristics

The respondent demographics generally matched those for Florida. Fifty-eight percent of the respondents were women. The average age of respondents was 44, with 92% White, 6% Black and 2% other. Ninety-six percent of the sample had completed high school, and 63% had some form of post-high-school education. Two-thirds of the sample were employed, primarily in the retail trade (25%) and services (22%). Forty-five percent of the sample had family incomes greater than $45 000.

4. Generator bias

All network generators we know of are subject to selection bias. Some bias is intended. For example, using the question 'Who are the people you can borrow money from?' is intentionally biased to select close alters. Close alters tend to be of similar age, race and gender as the respondent. We wanted a generator that had as little bias as possible.

Despite our efforts to reduce the bias associated with particular first names and alter characteristics, our analysis suggests that some bias still exists. Aside from the typical biases associated with large-scale surveys (i.e. sampling and response bias) there are two fundamental types of bias which are specific to this method.

Respondent selection bias is caused by certain types of respondents being excluded from the study because characteristics of the first names made it difficult to connect them to alters. A good example of this is the ethnically Asian respondent who knows mostly Asians. The list we provided would not generate many hits for this respondent since there were no names that were ethnically Asian.

Alter selection bias is also caused by the first names, but in this case we are concerned about the characteristics we infer to the personal networks being influenced by characteristics associated with particular first names. For instance, people named Helen tend to be older than people named Jennifer. People named John are males. Thus, our choice of first names will to some extent determine the characteristics of the network. In the following sections we will examine both respondent and alter bias extensively and suggest some methods for reducing their effects in future replications of this method.

5. Respondent selection bias

When analysing respondent selection bias it is important to think of the mechanisms by which such bias might be introduced. With first names there are three primary variables we should worry about: gender, ethnicity and age. Other types of respondent selection bias associated with first names are difficult to imagine, unless they are related to one of the three already mentioned. For example, will potential respondents who are

Table 1
Analysis of respondent selection bias by demographic variables

Variable	Did not complete survey	Completed survey	Significance level
Average age	45	44	0.336
Percent male	43	42	0.717
Percent black	12	6	0.001
Average years education	13	14	0.001
Annual household income > $45 000	34	45	0.001
Percent hispanic	11	5	0.001
Percent employed	59	67	0.001

heart surgeons select themselves out of the study because Johns, Helens, etc. are names not known to heart surgeons? This is clearly not the case. However, we *can* imagine respondents selecting themselves out of the survey based on gender, ethnicity or age.

Table 1 clearly shows that certain types of respondents were biased *against* completing the survey. Gender and age were not a significant factor with respect to respondent selection bias. That is, the list of names did not result in a disproportionate number of one gender or certain age groups finishing the survey. However, it is clear that Blacks and Hispanics were under-represented in the final sample. It also appears that low-income and lesser-educated respondents were somewhat under-represented. We find it difficult to imagine that the latter bias is due to the names used. Rather, we conclude that some people did not fully understand the method or were mistrustful of the legitimacy of the phone call and that these respondents tend to be less-educated which is strongly related to income levels. It is typical for telephone survey respondents to have higher incomes, on average, than in the target population.

Table 2 helps us analyze these problems further. This analysis was restricted to the first 1 027 respondents who completed the survey as the remaining 498 respondents were presented the first names in a different order to test for name-order bias. Table 2 presents results for the first 14 names presented to the 1 027 respondents. The table is restricted to these 14 names since they were presented to all respondents. Subsequent names would have a decreasing probability of being presented as respondents connected to alters from earlier names.

The first comparison to make from Table 2 is between the second two columns. Column 2 is an estimate from the Census sample of the actual percentage each first name represents of the US population. For example, 1.28% of all the people in the US are named Michael. The third column is the percent of the 1 027 respondents who knew someone with that name.

One indication of respondent selection bias with the first-name method would be a low association between the values in these two columns. If there were no first-name bias, that is no tendency for particular names to be associated with particular types of respondents, then we would expect a high degree of correlation between these numbers. In fact, the correlation *is* quite high at 0.86 ($p = 0.001$). Despite the biases associated

Table 2
Analysis of alter selection bias relative to age, race and gender

First name (in order read)	Percent of all first names in US	Percent of respondents who knew someone with name	Average age of respondents who knew someone with name	A	Percent of Black respondents who knew someone with name	B	Percent of male respondents who knew someone with name	C
Michael	1.28	69	44	<	65		72	
Mary	1.35	55	47	>	48		55	
David	1.15	57	44		40	<	61	
Sarah	0.26	33	45		28		31	
John	1.60	62	45		51	<	63	
Lisa	0.36	40	42	<	36		36	<
Chris	0.50	37	43	<	21	<	38	
Ann	0.19	39	46		40		36	
James	1.62	49	45		44		50	
Stephanie	0.20	28	42	<	36		26	
Robert	1.53	55	46		34	<	61	>
Deborah	0.25	37	43	<	34		34	<
Gregory	0.21	24	42	<	24		27	>
Helen	0.34	27	51	>	27		23	<

with some names, they do not cause a disproportionate representation of their frequency in the respondent's personal network.

Of course this does not mean there is no association bias. The fourth column shows the average age of those who knew someone with the name and the fifth column, labelled A, shows the results of a t-test between those who knew someone with the name and those who did not. The < indicates that the listed age is significantly lower than for those who did not know someone with the listed name; the > indicates the age is significantly older, and no sign indicates there was no statistical difference. For example, the name Helen demonstrates an age bias as the average age of those who knew someone with that name was 51, much higher than those who did not. Most names appear to be biased toward younger respondents which probably reflects the fact that we used a university registrar's list to create the name list.

As we mentioned, ethnic bias was one area we were concerned with when we designed our data collection instrument. We realized that it would be difficult to account for all ethnic groups with our list of names. Our choice was to try to create a list that represented most groups, or to attempt to create a list that represented them proportionally. For this research we chose the former strategy since we were constrained by the number of names we could effectively use in a telephone interview. As a trade-off for the increased representativeness of the sample, we chose common first names which would maximize the probability of a hit.

We admit that our choice of names was a problem for connecting with Asian alters. However, we do not believe that it was inherently biased for Blacks and Hispanics. Of course, there are examples of names that are exclusive to Blacks, such as Lashonda or

Latisha, but analysis of the registrar's list and other sources indicated that ethnically Black names were not the norm.

Note the percentages and tests in columns six and seven. We see that 65% of the Black respondents knew a Michael, a result not significantly different (by Chi-square) from Whites. In contrast, significantly fewer Blacks knew alters named David than did Whites. In all, four of the 14 names demonstrated a bias for Blacks. Our conclusion is that ethnically Black names, which became popular in the 1970s, are not a significant problem at this time, but may become a problem in the future. Ethnically Black names are much more common among Black women than among Black men, which is probably related to the same cultural trait which accounts for the diversity of women's names overall.

Bias for Hispanics was less of a problem, due in part to the tendency of Hispanics to translate Anglo names into their Spanish equivalents. Chi-square tests of the 1 027 respondents' recognition of the 14 names by Hispanic origin revealed only two names that presented problems for Hispanics (Ann and Helen).

Gender also presented difficulties with respect to bias. The last two columns show that some names are more likely to generate alter hits with males than females (Robert and Gregory) while others are biased in favour of female respondents (Lisa, Deborah and Helen).

What might we do to avoid respondent bias in the future? We could work to improve the explanation of the procedure to potential respondents. We would also suggest using the Census data to generate a list which is more representative of all ethnic groups. While it is logistically difficult to use more than 50 names, we could expose respondents to a wider list by randomly assigning respondents to different lists, or by randomly assigning names to each respondent so that each respondent gets a unique list drawn from a larger pool of names. Our survey software did not permit the latter option.

6. Alter selection bias

As with respondent selection bias we must think of the mechanism by which the characteristics of first names might affect the content of the part of the network we sample, that is, alter selection bias. Again we conclude that gender, ethnicity and age affects will be the primary problems. Whereas respondent selection bias is an all-or-nothing effect (either the respondent finished or not), alter selection bias can enter by degrees. This centers on our decision to eliminate from the sample anyone who did not connect with 14 alters, an arbitrary cut-off.

To explore the possibility of alter selection bias we first ran a series of ANOVAs on each first name and the various characteristics of alters elicited by the name. Our null hypothesis was that age, knowing level, duration of relation and frequency of contact would be the same between names. All of these variables showed significant differences ($p < 0.001$ for all variables except duration which had a $p < 0.05$) based on the first name. We must conclude that first names are associated with some attributes.

Unfortunately at this point the discussion becomes somewhat circular. On the one hand we do not want to determine the structure of a personal network by using cues (first names) which will bias the outcome. However, determining that the outcome is biased means that we forced alter selection disproportionate to that of the personal network of the respondent, which of course is different for each respondent.

In other words, making a list of unbiased cues assumes we know what the distribution of network characteristics should be so that we can construct the list to reflect that distribution. Yet the distribution varies by respondent; some have many older alters while others have few, some respondents know mostly women and others mostly men.

In the present study the most serious bias we faced was that of gender. While we are reasonably confident that the bias of specific names to particular age and ethnic groups was addressed to some extent by choosing relatively unbiased names, it was virtually impossible to do so for gender. Most names are either male or female. Although there are exceptions (e.g. Pat or Chris) we decided that there were not enough of these to make effective use of them, such as by using a list of *only* ambiguous names and letting the respondent choose the proportion of males and females that exist in their personal network. Rather, we decided to eliminate all gender-ambiguous names and use names that were gender-specific.

This presented another problem — How many male and female names should we include? Our research showed us that there is much more variety in the naming of girls than of boys. Indeed, the ten most popular male names account for 23% of all males in the US, while the top ten most popular female names account for only 11% of all females in the US. Deciding how many of each to include would influence the proportion of male and female alters represented in the respondent's personal network.

Ultimately we decided to alternate male and female first names, which due to the high variability of female names biased against getting accurate representation of female alters. Adding up the respective percentages in column 2 we see that the seven male names account for 7.89% of the US population while the female names account for only 2.95%, even though more than half of all US citizens are female.

What can we do about alter selection bias? Like respondent selection bias, the answer is to use large lists of first names that are randomized for each respondent and drawn from a large pool. Furthermore, the probability of drawing a name should be proportional to its frequency in the target population. Bias could be further reduced by stratifying this by frequency within regions, a possibility with the Census data.

7. Other types of bias

Another type of bias concerns the fact that we are allowing only one alter per name. Thus, a respondent may know three Roberts but we are allowing them to select only one. One reviewer suggested that the results should be adjusted to account for the disproportionate representation of common versus uncommon names. This could be done by weighting the results generated from Roberts more than for alters with the name Albert, for example. While this is an intriguing possibility, we decided to explore this in future replications of the study. As we mentioned above, this could also be addressed by using randomized lists and allowing names to occur more than once or with higher frequency across respondents if they are common names.

When working with lists of cues the possibility exists of some type of order effect where subsequent selections are based on criteria other than the presented cue. For example, if a respondent knew a Michael that was a strong tie, would there be a bias toward selecting subsequent alters who are also strong ties, perhaps by ignoring hits on weak ties until a connection to another strong tie was made. Brewer (in press) tested for order effects on tie strength based on these data and found no evidence of it.

8. Evaluating sampling efficiency

Despite the evidence of both respondent and alter selection bias, there is reason to believe that this method captures a far more representative sample of the personal network than other methods. Indeed, the high correlation between the percentage of the respondents who knew an alter with a given first name and the proportion of the US population with that name suggests that these biases were not fatal. We continue our assessment of this method by evaluating the sampling efficiency of the generator based on its findings.

There are several ways to evaluate the sampling efficiency of a network generator. The ideal would be to have a full census of the networks of a large group of people. We would then take representative samples of the same group and for each sample we would apply a different network sampling device, or name generator. Characteristics of each network sample would then be compared to similar characteristics of the network population. *Ceteris paribus*, a network generator that produced a sample more like the population would be preferred over a generator that produced a sample less like the population.

Lacking a network census, we can evaluate a network sampling device (in this case, the first-name method) in at least two ways: (1) compare the results to the results obtained with other methods; (2) assuming a representative sample of respondents, we can examine some characteristics of the alters in their networks and see if those characteristics make sense. We can also look for surprises — that is, for hints about features of network alters that merit further research.

9. Comparisons with other methods

Table 3 compares selected results from six studies. This table is an extension of one presented by Campbell and Lee (1991). They asked 690 residents of Nashville to list those neighbours with whom they had chatted or visited in the previous six months. (As a cueing device, the researchers presented respondents with neighbourhood maps.) Campbell and Lee compared results from their study with the results from three other studies of urban networks in North America, Wellman (1979), Fischer (1982), Marsden (1987). We have added results from our first-name study and from one of our earlier studies (McCarty, 1992).

Table 3
Comparison of six network generators (adapted from Campbell and Lee, 1991)

Generator characteristic	Wellman (1979)	Fischer (1982)	Marsden (1987)	Campbell and Lee (1991)	Free list Method	Current study
Generator cue	'People feel closest too'	Core network (several questions)	'People talk to about important matters'	Neighbours from map	Free list	First names of alters
Geographic area of respondents	Toronto, Canada	Northern California	U.S.	Nashville, TN	Gainesville, FL	Florida
Number of respondents	845	1050	1534	690	47	712
Number of alters named per respondent	4.7	18.5	3	14.7	60	14
Mean age of alters	-	37 urban 42 semi-rural	45	45	37	37
Mean education of alters	-	-	13	13	14	-
Percent of alters same Race as respondent	-	-	93	93	91	-
Percent of alters same Sex as respondent	-	58	58	58	57	53
Network density	0.33	0.44	0.57	0.52	0.27	0.36
Length of friendship	57% > 10 yr	urban – 14 yr rural -– 18 yr	6.4 yr	11.6 yr	10.4 yr	29% < 4 yr
Frequency of contact	150 contacts/yr	-	195 contacts/yr	117 contacts/yr	-	80 contacts/yr
Percent strong ties	-	38	-	32	27	26
Percent kin ties	50	42	55	-	25	29
Percent work-related ties	6	10	-	-	19	18
Percent neighbour ties	6	10	-	100	6	6
Percent other ties	38	38	-	-	50	47
Percent of alters living outside of respondent's city	25	33	-	-	58	53

Wellman (1979) asked 845 respondents from Toronto's East York district to list six 'persons outside your home that you feel closest too.' Marsden reports on the results of the General Social Survey (GSS) for 1986, which included questions on up to five alters with whom the 1 534 nationally selected respondents 'discussed important matters.' Wellman's East York study and the GSS focus on what we call *emotional support*.

Fischer (1982) elicited an average of 18 alters from 1 050 respondents in Northern California using a battery of questions designed to dredge the 'core network.' Some sample questions in Fischer's study include 'who would you leave your house with if you were going out of town?' and 'who could you borrow a large sum of money from?' The Fischer study focuses on what we call *social support*.

The four studies compared by Campbell and Lee were all based on large, representative samples of respondents, as was our first-name study. McCarty (1992), on the other hand, asked a non-representative sample of 47 respondents in Gainesville, Florida to simply list the first 60 people they could think of. The idea was to give respondents the opportunity to tell us about the kinds of people they thought were in their social networks rather than to ask respondents about specific kinds of network alters. We refer to this as the freelist study.

Despite the many differences in these studies, there are indications that the first-name method is a useful device for getting representative samples of networks and that it can be systematically improved.

1. The percentage of kin ties elicited by asking people whom they feel closest to and whom they talk to about important matters is around 55–60%. In any modern society, the number of non-kin will far exceed the number of kin in most people's networks. In a representative sample of networks, then, the percentage of kin ties elicited should be lower than in a sample focused on emotional support. Indeed, the percentage of kin in the freelist and first-name studies (25% and 29%, respectively) is around half that found in the East York and GSS studies. In an earlier study (Bernard et al., 1988) we asked respondents in several different cultures to name the starter in a small-world chain, given the location and occupation of the target. The proportion of kin alters in that study (including both blood relatives and affinals) ranged from 13% for the Florida sample to 15% for Paiute Indians, and 17% for Mormons in Utah and Ponapeans in Micronesia. In sum: studies of emotional and social support networks so far appear to contain about 55–60% kin; studies of networks based on getting an artificial task done contain about 15% kin; studies that try to sample across the range of network ties contain about 25–30% kin.
2. We expect a representative sample of the entire personal network to have a lower proportion of strong ties than would samples derived from methods that focus on core emotional and social support alters. Both the freelist and first-name methods produce more weak ties (26.5%, on average), than do the other methods for which we have data (35%, on average).
3. Weaker ties should be contacted less frequently than are stronger ties. Respondents in the first-name study reported an average of 80 contacts per year, about half the average (154) reported by respondents in the East York, Nashville and GSS studies.
4. In the East York and California studies, about 6% and 10% of alters were work related, respectively — about half the percentage found in the freelist and first-name

studies. These figures seem reasonable to us: we expect fewer work-related ties for alters who are part of the core emotional and social support network, and more work-related ties in a representative sample of network alters.

5. In today's highly mobile society, we expect many network ties to be with people who live outside one's own city; we expect, however, that core network ties will live in the same city as respondents. On average, in the East York and California studies, 29% of all ties were with people who lived outside the respondent's own city. In the freelist and first-name studies, this rises to 56%, on average — again, almost double.
6. Consistently across the California, Nashville and GSS studies, the sex of alters was the same as the sex of respondents in 58% of all cases. In the first-name study, men named 56% males, but women named just 46% females. The first-name method, then, appears artificially to lower the percentage of alters who are the same sex as the respondent. We see this, however, as evidence that the first-name method can be systematically improved to generate a representative sample of network alters.

10. What accounts for types of relations?

In the last section we compared the proportion of several relation types (kin, work, etc.) across six studies. We next examine what accounts for relation types across our respondents. We do this in two ways. First, we use characteristics of our respondents as independent variables and try to model the proportions of various types of alters. Second, we ask: what are the characteristics of respondents who tend to have more family relations compared to respondents who tend to have, say, more work relations?

11. Models

Recall that we asked respondents to tell us, in their own words, how they knew each of their 14 named alters and that we coded each relation as one from a list of 23 categories shown in Table 4. These categories were developed from in-depth interviews with respondents in several earlier studies (Bernard, 1982, McCarty, 1992).

Most of these relation types are obvious. For example, when asked how she knew an alter, if a respondent said that the alter was her mother, then this was coded as a 'blood relation.' If the respondent said that he knew the alter from church or from the Kiwanis Club, then the relation was coded as 'religion' or as 'hobby or organization.' If the respondent said 'oh, I've known him since we were children,' the relation was coded as 'childhood.'

Some of the categories are not so obvious. If the respondent said that an alter was her 'boyfriend's sister,' then that alter was coded as a 'step relation' — that is, through another person. We have found that step relations make up a substantial part of all networks (Bernard, 1982) and that some step relations are quite complex. One respondent described an alter as her 'ex-boyfriend's ex-girlfriend'; another said 'he's the drummer in my friend's band.'

Table 4
Distribution of relations and results from regressions on respondent characteristics

Coefficients (significance) for significant respondent characteristics [a]

Relation type	Average proportion	Age	Gender	College	Income > \$45 000	Employed	Black	Republican	Married	Model R^2
Family by blood	0.19				−0.03 (0.02)	−0.04 (0.01)			0.03 (0.01)	0.06
Family by marriage	0.11	0.0007 (0.01)	0.03 (0.01)	−0.05 (0.01)		−0.03 (0.01)			0.06 (0.01)	0.19
Work	0.18					0.18 (0.01)		0.03 (0.04)		0.17
Situational	0.08		−0.03 (0.01)						−0.03 (0.01)	0.02
Step	0.19					−0.04 (0.02)	−0.06 (0.02)	−0.03 (0.01)	−0.03 (0.01)	0.03
School	0.07	−0.002 (0.01)		0.04 (0.01)		−0.02 (0.05)	0.05 (0.01)		−0.03 (0.01)	0.17
Religion	0.04							0.02 (0.1)		0.02
Neighbour	0.06	0.001 (0.01)				−0.03 (0.01)				0.10
Hobby or Organization	0.04									0.00

[a] All models, except for Hobby and Organization, are significant at 0.01. The order of characteristics in the table reflects the order in the model.

We coined the term 'situational relation' because respondents often tell us that their relation to a particular alter is 'just because of the situation' or 'because we happen to be in the same place a lot.' While this category of relation is true to our respondents' ideas about how they are connected to various alters, the category does pose some coding problems. For example, the relation between an accountant and a janitor at her office could be coded as a work relation or a situational relation. The coding of this was left to the interviewers who were allowed in this study to code for only one type of relation per alter. Further tests of the first-name method should allow for coding of more than one relation at a time, although this presents problems for analysis.

If seven of a respondent's 14 alters were kin, then that respondent's kin were 50% of his or her sampled alters. (We infer, then, that 50% of all their alters are kin.) Nine of the 23 categories of relations accounted for 96% of all relations as coded by interviewers in this survey. Column 2 of Table 4 shows the average proportion of these nine relation types among the alters of respondents ($n = 712$) who completed the first-name task in our survey. For example, as shown in Table 4, on average 19% of the alters of our respondents were blood relations.

Reading across Table 4 shows the results of models with respondent characteristics as independent variables and type of relation as dependent variables. Only four of the models (for relations based on affinal kinship, on being a neighbour, or through work or school) explain at least 10% of the variance in the proportion of particular relations. The good news, however, is that there are few surprises, either, which increases our confidence that the first-name method dredged up a representative sample of network alters.

For example, Table 4 shows that, on average, 11% of our respondents' relations were family by marriage. A respondent's age and gender, along with their employment status, marital status and whether or not they went to college account for 19% of the variance in the proportion of affinal relatives among our respondents. Now, whether or not the respondent is married accounts for 8.5% by itself. The positive association with age, and the negative association with being employed or having a college education suggests that those with many marital relations tend to be housewives who do not work outside the home — just as we would expect.

Also as we expect, work relations are explained most by the employment status of the respondent. Work relations are weakly explained by whether or not the respondent is affiliated with the Republican party. This *was* a surprise, so to test the possibility that party identification was a proxy for employment, we introduced an interactive term combining the two and recalculated the model. This term increased the significance of all terms, and raised the parameter estimate for political identification. Republicans, despite their employment status, have significantly higher proportions of work relations.

Relations based on having gone to school with an alter are, as expected, associated positively with having a college education and negatively with age, being employed and being married. In other words, respondents with a high proportion of relations based on their school are young, single and recently graduated. An unexpected finding is that Black respondents have 2.5 times the proportion of school relations as do non-Blacks (15% compared to 6%). This may be due to the fact that the first-name list is biased toward ethnically White names and that Blacks tend to know Whites through school

Table 5
Characteristics of respondents and their network when one relation type predominates in the personal network

Characteristic (relations type)	All	Blood	Marital	Work	Situational	Step	School	Religious	Neighbour	Hobby/Org.	Multiplex
Number of respondents	712	123	40	154	53	125	34	25	12	17	129
Mean respondent age	44	49	55	40	42	44	33	52	56	48	44
Male respondents (%)	42	35	32	50	60	34	38	36	50	47	43
Mean education in years	14.1	13.9	12.7	14.6	13.6	13.9	15.4	14.1	15.3	13.7	14.1
Mean household income	$53 500	$47 800	$39 200	$56 000	$57 400	$60 600	$45 800	$65 300	$46 400	$58 500	$55 500
Employed respondents (%)	67	51	30	94	77	61	79	44	25	47	68
Black respondents (%)	06	08	05	07	06	03	15	04		0	05
Republican respondents (%)	35	38	35	39	28	25	26	60	58	47	37
Married respondents (%)	57	69	73	53	42	51	26	80	58	53	61
Mean network density	0.36	0.47	0.43	0.35	0.33	0.32	0.35	0.32	0.35	0.29	0.33
Mean age of alters	37	39	41	36	36	37	30	39	48	42	37
Alters living outside respondent's city (%)	53	60	64	53	45	51	42	44	46	53	54
Mean level of knowing, 1–5	3.2	3.5	3.4	3.1	3.1	3.2	3.3	3.1	3.3	2.9	3.1
Mean level of information, 1–5	3.1	3.4	3.3	3.0	3.0	3.1	3.2	2.9	3.2	2.7	3.0
Alters known for less than 4 yr (%)	29	16	12	40	39	29	30	29	24	31	30
Mean contact days per year	80	64	64	112	96	60	81	59	82	57	76
Mean network size	432	299	327	445	480	531	336	354	416	384	485

more than other relation types. Although the sample of Black respondents is significantly ($p = 0.01$) younger than are non-Black respondents (by nearly a decade), race is still a significant factor in the model after controlling for age.

On average, 6% of respondents' alters were their neighbours. This relation is positively associated with age and negatively with being employed. These findings are quite sensible. Those age 65 and over have twice as many neighbour relations (10%) as do those under 65 (5%). Those who are employed have high proportions of work relations and are home less of the time.

12. Whose networks are mostly based on certain relations?

We next looked at those respondents whose proportions of blood relations, marital relations, work relations, etc. were higher than any other type of relation. Some respondents had two or more relation types tied for first place, while 583 had a predominant relation type — that is, they did not have equal proportions of two kinds of predominant relations. Although this analysis forces us to exclude some respondents, i.e. those who do not have a dominant relation type, we see it as a worthwhile sacrifice given that most respondents do have a dominant category and evince different characteristics.

Table 5 presents a summary of the data on these 583 respondents. We have no reason to suppose that these 583 respondents are different from the full set of 712. The 583 are nearly identical to the 712 on their average age, income and education, the proportion employed, the age of their alters, and so on.

At the bottom of Table 5 we have added data on the estimated size of the networks for various types of respondents. This was estimated using a method detailed in Killworth et al. (in press). In brief, estimates by respondents of the number of alters they know in various subgroups of known size are used to scale up to estimates of network size.

We examine the data in Table 5 for hints that the first-name method produces sensible results and for hints about network composition that merit further research.

13. Blood relations

The data on blood-related alters offer strong support for the efficiency of the first-name generator. Sixty-five percent of the 583 respondents who named more blood relations than any other type are women and their network density is the highest of all categories — that is, blood-related alters are more likely to know one another than are, say, the people known from work.

Furthermore, respondents who have many blood-related alters report knowing more about their alters than do other respondents. This seems very sensible. Respondents in this category also have the smallest estimated network size. This may be evidence that reliance on strong ties produces attenuated networks, and conversely. This certainly demands more study.

14. Marital relations

Unsurprisingly, respondents in this category are very much like those who name more blood relations, though a *t*-test shows that respondents who name more marital relations have significantly ($p = 0.01$) less education. Only those who report more relations with neighbours are more likely to be unemployed. We think this is evidence that: (a) housewives who do not work outside the home rely heavily on relations by marriage, and (b) people who rely more on neighbours for network ties are retired (their average age is 56).

Average network density is relatively high, again reflecting the expected ties among in-laws. The proportion of alters living outside the respondent's city is very high and is comparable to the figure for those who have more blood relations. This reinforces the common view that respondents move for reasons other than to be near their family.

It also shows that, for about a fifth of the population in Florida, at least, having a large part of their relatives out of town does not diminish the strength of their ties to those relations. The number of contact days is low (which fits with the high proportion of out-of-town ties), but both measures of tie strength (the knowing and information scales) are highest for the marital- and the blood-relation categories. Of course, for both categories, the proportion of alters known for less than four years is far lower than for any other category.

Note that 73% of those who name more network ties with affinal relatives are married. This means that 27% of those in this category are *not* married. These 11 people are either widowed or divorced but continue to have strong ties with their affinal relatives. With only 40 people in the marital-relation category, these data only hint at a phenomenon that requires further exploration.

15. Work relations

More people (154 out of 583) name work as the dominant source of their alters. Respondents in this category are significantly ($p = 0.01$) younger than those who name more relatives (by blood or by marriage), are more educated ($p = 0.01$) and have higher ($p = 0.01$) incomes. Unsurprisingly, 94% of the people in this category are employed, but we would like to know more about the circumstances of those nine people (6%) who are not. The high proportion (40%) of relations reported to be of short duration and of frequent contact (112 days per year) seem typical of the work environment.

16. Situational relations

Many social ties result from simple circumstance. We meet and get to know the names of mail carriers and of parents whose children attend the same school as do our own children — without ever forming more than the weakest of ties. Note, however that the measures of average tie strength (knowing and information) and income are identical for this group and for the group whose ties are based more on work, and that a high proportion (77%) of this group is employed. We think this is evidence that some respondents perceive work relations as just circumstantial and we intend to examine this further.

17. Step relations

Respondents with primarily step relations have the largest network size, and among the lowest number of reported contact days per year. These are respondents who have access to many alters but those ties depend on gatekeepers. Most members of the step-relations group are women whose household incomes are above the average for all respondents. Density is among the lowest of all groups.

18. School relations

Predictably, respondents with many school relations are the youngest and most educated of all groups. Their incomes are below average and 74% are unmarried (predictable from their average age) but a high proportion (79%) is employed (predictable from their high education). Unsurprisingly, the mean age (30) of the alters for this group is the lowest of any group as is the proportion who live outside the same city.

19. Religion relations

The 25 respondents whose networks are based more on relations associated somehow with their religion are older than most respondents and have the highest average incomes overall. Eighty percent are married and 60% are affiliated with the Republican party.

This group reports the lowest mean number of contact days per year (59) with their alters — about what might be expected from weekly and holiday attendance at church. With low frequency of contact, mean network density is also low, as are both measures of tie strength. For a small percentage of people in Florida, then (25 of the 583 in this part of the study), infrequently seen, weakly known ties through their religious affiliation are nevertheless most salient.

20. Neighbor relations

Respondents associated mostly with neighbor relations are the oldest and among the most educated. Seventy-five percent are unemployed and the mean age of their alters is the highest of any group. We conclude that this small group of just 12 respondents are mostly retired people. These 12 respondents are just 2% of the 583 in this part of the study. A small percentage of elderly, retired people, then, rely a lot on their neighbours.

21. Hobby relations

Like the respondents in the religion group, the 17 people who mainly named hobby- or organization-based alters report a low number of contacts per year (57). Like church attendance, this suggests weekly meetings. The strength of tie measures are the lowest for any category suggesting superficial relations. Network density, as expected, is the lowest of any group.

22. Multiplex relations

Some respondents did not have one relation type that dominated the others. We have labelled these respondents' networks as multiplex. For this study we define a multiplex network as one where two or more relation types are tied for the maximum proportion of all relation types. For example, a respondent who has 29% blood relations and 29% work relations would have a multiplex network, in contrast to the respondent who has 36% blood relations and that is the highest proportion of any relation type.

There were no clear patterns of relation type combinations that made up the multiplex group. Some involved blood and work relations, others involved work and step relations. There was a sufficient mix to justify the conclusion that multiplex relations, as we have defined them, involve many different relation types.

Because the multiplex networks tend to involve many different relation types, their characteristics as shown in Table 5 are most like the average for the entire sample of 712. Unlike most of the networks dominated by one relation type, none of the characteristics of the multiplex networks diverge dramatically from the characteristics of the whole. They appear to be a representative subsample.

23. Conclusions

Despite the selection biases, the first-name method appears to be a useful and efficient tool for collecting large amounts of data quickly and at relatively low cost about the total social networks of a population. The method can be used in telephone surveys. Researchers need only vary the information collected about network alters to apply this method to a variety of substantive areas, such as consumer behaviour or medical decision-making. The method produces associations with respondent and network characteristics that are sensible, and in many cases fully expected.

It remains the case that many studies of networks require focused, not representative samples of alters. A study of social support among the elderly in a retirement home, for example, might not be based on a representative sample of alters. In contrast, a study of how network alters affect the choice of a physician might usefully be based on a sample of alters from the total network.

Furthermore, the method clearly needs additional testing. We do not know how best to select a list of first names so that the list is appropriate to particular populations with particular ethnic mixes. Fourteen alters may be too few to test some network hypotheses, and 50 names may be too short a list with which to cue respondents.

In further tests of the method we will try to improve our results. Fortunately, the U.S. Census Bureau has recently released a list of virtually all first names in the United States, and the percentage of each first name in the population. We will not have to rely, then, on university registrar lists or other lists that may contain unknown biases. Researchers who want to use the Census Bureau list will find it on the World Wide Web at: http://www.census.gov:80/genealogy/names/

References

H.R. Bernard, INDEX: A respondent-defined experiment in social structure, *Social Forces* 61 (1982) 99–133.

H.R. Bernard, P.D. Killworth, M.J. Evans, C. McCarty and G.A. Shelley, Studying social relations cross-culturally, *Ethnology* 27 (1988) 155–79.

H.R. Bernard, E.C. Johnson, P.D. Killworth, C. McCarty, S. Robinson and G.A. Shelley, Estimating the size of an average personal network and of an event subpopulation, in M. Kochen (ed.), *In The Small World* (Ablex Publishing Corporation, Norwood, NJ, 1989).

D.D. Brewer, No associative biases in the first name cued recall procedure for eliciting personal networks, *Social Networks*, in press.

Karen E. Campbell and Barrett A. Lee, Name generators in surveys of personal networks, *Social Networks* 13 (1991) 203–221.

Claude S. Fischer, *To Dwell Among Friends: Personal Networks in Town and City* (The University of Chicago Press, Chicago, IL, 1982).

Linton C. Freeman and Claire R. Thompson, Estimating acquaintanceship volume, in M. Kochen, *In The Small World* (Ablex Publishing Corp, Norwood, NJ, 1989).

Mark Granovetter, *Getting a Job: A Study of Contacts and Careers* (Harvard University Press, Cambridge, MA, 1974).

Eugene Johnsen, H. Russell Bernard, Peter D. Killworth, Oene A. Shelley and Christopher McCarty, A social network approach to corroborating the number of AIDS/HIV+ victims in the US, *Social Networks* 17 (1995) 169–187.

P.D. Killworth, H.R. Bernard and Christopher McCarty, Measuring Patterns of Acquaintanceship, *Current Anthropology* 23 (1984) 318–97.

P.D. Killworth, E.C. Johnsen, H.R. Bernard, G.A. Shelley and C. McCarty, Estimating the size of personal networks, *Social Networks* 12 (1990) 289–312.

P.D. Killworth, C. McCarty, E.C. Johnsen, G.A. Shelley and H.R. Bernard, A social networks approach to estimating seroprevalence in the United States, *Social Networks*, in press.

Edward O. Laumann, John H. Gagnon, Stuart Michaels, Robert T. Michael and L. Philip Schumm, Monitoring AIDS and Other Rare Population Events: A Network Approach, *Journal of Health and Social Behaviour* 34 (1993) 7–22.

Nancy Howell Lee, *The Search for an Abortionist* (University of Chicago Press, Chicago, IL, 1969).

Peter V. Marsden, Core discussion networks of Americans, *American Sociological Review* 52 (1987) 122–131.

Christopher McCarty, Perceived clique definition in ego-centered networks, *Ph.D. Dissertation* (University of Florida, Gainesville, FL, 1992).

I.S. Pool and M. Kochen, Contacts and influence, *Social Networks* 1 (1978) 5–51.

Barry Wellman, The community question: The intimate networks of East Yorkers, *American Journal of Sociology* 84 (1979) 1201–1231.

Holly A. Williams, Social support, social networks and coping of parents of children with cancer: Comparing White and African American parents, *Ph.D. Dissertation* (University of Florida, Gainesville, FL, 1995).