Title
An inductive typology of egocentric networks

Authors
Giannella, E
Fischer, CS

Publication Date
2016-10-01

DOI
10.1016/j.socnet.2016.02.003

Peer reviewed

Social Networks 47 (2016) 15–23

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

An inductive typology of egocentric networks

Eric Giannella*, Claude S. Fischer

University of California, Berkeley, 410 Barrows Hall, Berkeley, CA 94720-1980, United States

ARTICLE INFO

Article history:

Keywords:
Personal networks
Egocentric networks
Clustering
Random forests
Typology

ABSTRACT

We apply Random Forests to detailed survey data of social relations in order to derive an inductive typology of egocentric networks. Beginning with over 40 descriptors of 1050 northern California respondents' networks, we combine 21 of these into seven dimensions, the extent to which those networks display: (1) interaction with nonkin, (2) proximity to kin, (3) overall involvement with kin (including support), (4) support from nonkin, and the extent to which (5) church, (6) work and (7) extra-curricular activities shape connections with others. We use these dimensions to reliably place 985 of the 1050 observations into types: career-and-friends (24%), family-and-community (20%), family-only (16%), untethered (8%), energetic (7%), withdrawn (6%), and home-and-church (5%). In the second part of the analysis, we describe the social and demographic attributes of respondents that predict membership in each cluster to present a richer picture of the network typology, as well as to confirm that the types have face validity.

Identifying a typology of egocentric networks has been at least an implicit task in social network analysis since its early years. Researchers have distinguished types of egocentric networks by their low versus high degree of mesh (Barnes, 1954), close- versus loose-knittedness (Bott, 1955), local versus cosmopolitan orientation (Merton, 1968: Ch. 12), with or without "weak ties" (Granovetter, 1973), kin- versus friend-based (Wellman, 1979, p. 1211), and so on. Later efforts, which we review below, developed more complex network typologies more systematically and often in larger data sets. Our contribution in this article furthers this work, trying to answer the question, What are the basic types of egocentric networks?

We drew upon over 40 descriptors of networks from a survey unusually rich in information about over 1000 respondents' many personal ties. And we deployed a technique new to sociologists for grouping observations in order to extract an inductive typology. In the first part of our analysis, we find that the respondents' networks are best distinguished by seven dimensions, the extent to which those networks display: (1) interaction with nonkin, (2) proximity to kin, (3) overall involvement with and support from kin, (4) support from nonkin, and the extent to which (5) church, (6) work and (7) extra-curricular activities shape connections with others. We then use these seven dimensions to place 985 of our 1050 respondents into typological clusters. We name clusters with more than 5% of observations: *Career-and-Friends* (24%), *Family-and-Community* (20%), *Family-Only* (16%), *Untethered* (8%), *Energetic* (7%), *Withdrawn* (6%), and *Home-and-Church* (5%).[1]

In the second part of the analysis, we describe the social and demographic attributes of respondents that predict membership in each cluster. For instance, we find that being younger, more educated, and childless predicts having networks of the *Career-and-Friends* type, while being married, a parent, and having a declared Christian affiliation predict membership in the *Family-and-Community* type. These secondary analyses allow us to present a richer picture of the network typology, as well as to confirm that the types have face validity.

We use the Northern California Community Study (NCCS, PI: Claude Fischer; ICPSR #07744) which describes in great detail over 19,000 ties in the egocentric networks of over 1000 respondents living in that region in 1977–1978. These data have been previously analyzed (Fischer, 1982; Feld, 1984; Blum, 1985; Rook, 1987; Marks, 1994), but not to our knowledge for this purpose. The NCCS survey has the disadvantages of being regional and almost 40 years old. Its age, in particular, means that the role of post-1980s communications technology is not assessed. Nonetheless, like other venerable data sets that have been repeatedly used by network analysts (e.g., the 1950s medical innovation research (Coleman et al., 1957); 1970s karate club data (Zachary, 1977)), it has continuing

* Corresponding author.
E-mail addresses: giannella@berkeley.edu (E. Giannella), fischer1@berkeley.edu (C.S. Fischer).

[1] The remaining four clusters were: *Semi-Isolated* (~4.5%), *Nonkin-as-Kin* (3%), *Sociable* (3%) and *Just Activities* (3%).

http://dx.doi.org/10.1016/j.socnet.2016.02.003
0378-8733/

value. In addition, the NCCS contains an unusually high number of network descriptors for a survey.

1. Extracting types of egocentric networks

There are many reasons social scientists might wish to identify the basic dimensions and types of social networks, notably the proposition that network patterns themselves, above the level of dyads, affects egos' lives (Wellman and Gulia, 1999; for an application to medical care, see Pescosolido et al., 1998). In a review of clustering over the past five decades, Jain identifies three basic goals of researchers:

> *Underlying structure:* to gain insight into data, generate hypotheses, detect anomalies, and identify salient features.
>
> *Natural classification:* to identify the degree of similarity among forms or organisms (phylogenetic relationship).
>
> *Compression:* as a method for organizing the data and summarizing it through cluster prototypes (Jain, 2010: 653, italics original).

Our goals are the "underlying structure" and "compression" of personal networks. In trying to characterize personal networks, researchers typically use single measures, such as the number of people from whom ego could borrow money, although we intuitively and empirically know that these measures are correlated, for example, with ego's closeness to his or her relatives. Having a sense of the "packages" of network attributes allows us to more efficiently distinguish individuals by their networks.

Similarly, it is also difficult to know which network features are more central than others. For instance, does the number of *nonkin* respondents socialize with better predict other network characteristics than does the number of *kin* they socialize with? The process of deriving a typology may help us both discover which network attributes go together and which are most important in distinguishing personal networks.

A second motivation pertains to comparison across time or society. Typologies would permit comparison of each society's distinctive formation of personal networks. For instance, changes in gender dynamics in recent decades in the United States suggest that we should observe convergence in network types that apply to men and women (see Smith et al., 2014).

Third and most abstractly, a robust typology can uncover some of the underlying "rules" or "forces," whether ecological constraints or cultural expectations, that influence individuals' network construction. Such conditions make some configurations of personal networks more likely than others. Our results showed that respondents' levels of kin sociability and of practical kin support might as well be treated as one dimension, because receiving some kin support almost always coincided with seeing kin socially. In contrast, support from nonkin did not consistently accompany moderate or even high social engagement with nonkin.[2]

Previous efforts at identifying typologies have been instructive, yet constrained by the relatively small number of relational characteristics that they use. Antonucci et al. (2010) summarize many previous efforts to extract, via some form of clustering procedure, typologies of egocentric networks: "Across studies... several relatively robust network types can be identified... specifically, a diverse or diffuse network type, a restricted or socially-isolated network type, a friend- or community-focused (or both) network type, and a family-focused network type.... The diverse network type tends to be the most common, whereas the family-focused and restricted network types tend to be the least common...." (p. 459). Our findings are consistent with but substantially more refined than this summary.

In general, the data for early studies consisted of respondents providing global descriptions of their networks. Surveys ask respondents to describe attributes such as type of support, reliability of support, contact frequency, and felt closeness that they have with pre-determined *categories* of ties, such as parents, children, and "friends" (e.g., Birditt and Antonucci, 2007; Fiori et al., 2006). Alternatively, many studies ask respondents about the kinds of support they get, such as childcare help, or money, and what *kinds* of alters provide them, kinds such as kin in the household, relatives outside the home, and neighbors (e.g., Gibney and McGovern, 2011). Previous studies are limited in a few other, key ways, as well. The samples are often of special strata, particularly the elderly. Their surveys generally use network-level measures, which can make it more difficult to spot differences between networks, such as those with confidants who are kin versus nonkin. Typologies based on these data therefore rely upon more aggregated data. (Two modest exceptions are Hennig (2007) and Kim (2012).[3])

The NCCS data allow researchers to build up rich descriptions of egocentric networks from complex, detailed descriptions of specific relationships. For example, the data describe whether ego gets emotional, financial, practical, or other kinds of support from each one of his or her "friends," not just general support from the category of friends. As described below, we generate over 40 variables to be used for clustering, almost all of them derived from multidimensional descriptions of many ties per respondent. For example, the NCCS respondents named a median of 17 alters, a median of two with whom they discussed "personal matters," and a median of seven with whom they had socialized in the prior three months. This is a degree of fine-grained data largely unavailable in the prior efforts to create typologies, at least in American data.

2. Data and methods

The NCCS surveyed, in person, 1050 people living within a roughly 200-mile radius ranging north and east of San Francisco, including the city itself. Interviewers asked respondents to name, in almost a dozen name-eliciting questions, people in their personal networks. As detailed in the appendix, respondents named those with whom they spent social time, depended upon for advice, received practical assistance, would ask for a loan, and so on. Interviewers then compiled those names and asked respondents to describe the alters, such as their genders, and the nature of the dyadic ties, such as how nominees were connected to and how far they lived from the respondents. The survey instrument and a detailed description of the methods can be found in the appendices to Fischer (1982).

[2] Findings about "configurations" such as these can be compared to research on personal networks in other cultures. One example that serves as an immediate contrast to the kin/nonkin distinction we draw is Bastani's work on personal networks in Iran (2007), in which she finds that there is a certain guarantee that kin will be around – particularly for the old.

[3] Hennig applies Wellman's (1979) typology of the "lost", "saved" and "liberated" to extensive 2003 German survey data. She first finds three variable clusters: neighborhood embeddedness (including proportion of network in neighborhood, proportion of family in neighborhood, proportion of network members contacted multiple times per week); proportion kin (proportion relatives in network, density); range (network size, multistranded ties)). She maps these to Wellman's typology and finds: "None of the German models show a loss of community: there are relatives, friends, neighbors, and acquaintances in all three of the cluster centers. The difference lies in the proportions" (388). Kim (2012) uses the GSS's "personal matters" question (disputed, e.g. Bailey and Marsden, 1999; Bearman and Parigi, 2004; Small, 2013) to create typologies based on the descriptions of the up-to-five names respondents to the National Social Life, Health, and Aging Project survey provide. He distinguishes, for example, spouse plus children networks from spouse-only from spouse plus others networks.

2.1. Analytical strategy

We first provide a one paragraph overview of our methods for the reader with less interest in our procedure. In the rest of the section, we go on to describe Random Forests and the essential steps of our procedure. More details can be found in the Supplementary data.

We began by identifying 43 variables from the hundreds of potential descriptors in the NCCS survey that might help to characterize personal networks. We used Random Forests (Breiman, 2001) to create similarity measures among all the observations in our dataset using these 43 variables. We clustered the observations using the derived similarity measures. We then ran Random Forests a second time to assess how accurately individual networks could be classified into their respective clusters. We discovered that we needed to combine variables into "dimensions" of personal networks to reliably cluster a high proportion of observations. To do so, we treated the highly predictable clusters (containing 55% of observations) as more informative than the less predictable clusters and, using these distinctive cases, combined variables into substantively meaningful "dimensions" on the basis of their correlation and covariance. That is, we first used clustering to discover the key, discriminating measures; 21 out of the initial 43 emerged. We then constructed 7 "dimensions" from those 21 most discriminating measures, which we treated as a means of efficiently describing the "space" of personal networks in our data. We used the 7 "dimensions" to create similarity measures with Random Forests, clustered the observations, and identified 11 relatively coherent groups (placing 985 of our 1050 cases with about 88% accuracy) as a typology of egocentric networks.

2.1.1. An overview of random forests

We treated the construction of the typology as an exercise in parsimony; at each step we tried to synthesize as much as possible while discarding as little information as possible. We relied heavily on a relatively new technique, Random Forests (henceforth, RFs), at each step. Compared to traditional clustering techniques, such as latent class analysis (LCA), RFs can better sort through a larger number of variables to identify which combinations best characterize particular subsets of observations. They are also insensitive to differences in variable scales and can use binary and categorical variables alongside continuous ones. For the past decade and a half, RFs have gained wide adoption in other disciplines, such as computer science, genomics, and medicine, in situations where there are many potential variables that might be useful for clustering observations (Shi et al., 2005; Genuer et al., 2010). RFs have also been shown to be robust to outliers, to inclusion of extraneous variables, and to multicollinearity (Cutler et al., 2012). RFs excel in aiding researchers to discover which variables are most discriminating in very large datasets containing complicated interactions. Another benefit of RFs is that they perform very well in classification tasks (Diaz-Uriarte and Andres 2006; Caruana and Niculescu-Mizel, 2006). We use their classification accuracy to select the optimal number of clusters at each step.

RFs are an ensemble method: they combine the results of many varied analyses. Each tree in the RF is given a random subset of data to classify (a training sample). These observations begin at the "top" of the tree and are sorted toward their respective classifications using bifurcations based on variable values. At each bifurcation, the variable (and variable value) that best divides the observations into the correct classifications is chosen from a randomly sampled subset of variables. The trees are grown until they reach a specified terminal node size. Fitted based on the random training data, each tree is then used to predict the membership of observations that were not in the random subsample, thereby providing a prediction error for each tree. One benefit of using randomly sampled data for each tree and randomly sampled variables at each split is that it allows for identifying variables that are particularly helpful in describing subsets of observations.

Absent any pre-existing classifications, RFs filter observations down many trees in order to see which observations frequently end up in the same terminal nodes. To do this, the algorithm simulates data in order to initiate a mock classification exercise in which it tries to distinguish between real and simulated observations. The purpose of this exercise is to see which *real observations* end up together in the same terminal nodes. Observations that filter down to the terminal node of a tree together can be taken to be more similar, because they have survived several splits from other observations on the basis of key variable values. Because the data for each tree represent a random subsample of the entire dataset and the variables evaluated for each split are a random subsample of all variables, key similarities among observations that may be hidden in some trees might emerge in others. Each of the hundreds or thousands of trees generated by RFs thereby represents a new opportunity to learn about similarities among observations. Overall similarity scores among observations can be calculated using the frequency with which they appear together in terminal nodes across all the trees in a forest. The RF procedure outputs an $n \times n$ matrix with these similarity scores, which can be used for clustering.

In addition to providing information about the *similarities among observations*, we used RFs to obtain *relationships among variables*. Trees that more accurately classify observations identify the more useful variables for making those classifications. Variables can also be swapped out to measure how important they are to classification accuracy. RFs can output variable importance scores based on an entire dataset or based on particular groups within the data (e.g., for classifying observations into their respective "types").

2.1.2. Specific procedure

While more details appear in the methodological appendix, we present a brief sketch here. We first identified all the variables from the NCCS dataset that might pertain to key aspects of personal networks that have been researched in previous scholarship. Specifically, from the many variables available in the NCCS, we sought out those that dealt with degree of support, distance to kin and nonkin, homophily, opportunities for meeting alters, social styles, and tie strength. We identified 43 non-redundant variables that pertained to these categories. With these variables, we used Random Forests (*Salford Predictive Modeler 7.0*) to create similarity measures among all the observations in our dataset. We used the similarity matrix to cluster observations with a hierarchical clustering algorithm (*R* package *cluster*, function *agnes*). We chose the number of clusters that maximized the predictive power of RFs in classifying observations into their respective clusters (R package *randomForest*; Fig. A1 in the Supplementary data).

Starting from this initial set of clustering results, we combined key variables into "dimensions" of personal networks. We determined which variables to combine using the *R* package *clustOfVar*, which allows both categorical and continuous variables (Chavent et al., 2012). We treated the more predictable clusters as more informative about the discriminatory power of variables than the less predictable clusters. Thus, we only provided *clustOfVar* with the cases from clusters with a predictive accuracy of better than 75% (14 of the 35 clusters, comprising 579 of the 1050 observations, yielded better than 75% accuracy).

We chose the optimal number of groups of variables relying upon both our own substantive interpretation of the variable groupings as well as an adjusted Rand criterion provided within the *clustOfVar* package (Fig. A2 in Supplementary data). In addition, we tried to keep variables that were ranked important by the RF algorithm (Table A2 in Supplementary data). We discarded variable groups that could not be substantively interpreted or that contained only one variable (see Appendix for detailed discussion). In the end, we kept 7 of the 14 composite variables generated by the *clustOfVar* clustering procedure; these 7 jointly combined 21 of the 43 original variables. We take these 7 dimensions to efficiently describe the "space" within which egocentric networks can be placed.

Combining the key variables so identified into "dimensions" offers two benefits. First, combining variables allows more individuals to be reliably placed into groups or "types." Second, the reduction of variables into fewer dimensions allows the differences between the resulting types to be more interpretable. Often the combinations joined variables that were complementary and not simply substitutes for one another. For example, the "Nonkin Support" dimension includes a broad measure of support from nonkin based upon the number of nonkin whom the respondent relied upon to discuss personal matters, obtain advice, and get an emergency loan.

With these 7 composite variables (rather than the original 43), we once again used RFs to create similarity measures among all the observations in our dataset. We again clustered the observations using hierarchical clustering and chose the number of clusters that maximized the predictive accuracy RFs could achieve (see Fig. A3 in Supplementary data). The optimal number of clusters appeared to be 18, but we only consider 11 of those clusters to be personal network "types" (*n* obs. = 985). The mean predictive accuracy for the 11 clusters we present is 88%, meaning that roughly 8 out of 9 times Random Forests is able to learn enough from the variables to place an observation into its "correct" type. The 7 clusters we do not consider "types" and do not present are very small (65 cases in total) and hard to predict. Trimming them off from larger clusters increased the coherence of those 11 clusters.

Our process can be summarized in more discrete steps:

1. Choose variables that may represent important aspects of personal networks (43 in all).
2. Use Random Forests to create similarity measures among all 1050 observations (*SPM 7*).
3. Cluster the observations using these similarity measures into sets of 10–50 clusters (*R* pkg. *cluster*, function *agnes*).
4. Use Random Forests a second time to learn the characteristics of each cluster and predict which observations belong to which clusters in order to choose the best number of clusters (*R* pkg. *randomForest*).
5. Treat the most "coherent" clusters (prediction error < 25) as being more informative about underlying patterns in the data (*n* observations = 579).
6. Cluster variables using the 579 observations from these "coherent" clusters (*R* pkg. *clustOfVar*).
7. Based on the results from the previous step, combine 21 of the 43 original variables into 7 composite variables.
8. With the 7 composite variables, use Random Forests a third time to create similarity measures between all 1050 observations (*SPM 7*).
9. Cluster the observations using these similarity measures into sets of between 2 and 20 clusters (*R* pkg. *cluster*, function *agnes*).
10. Use Random Forests to learn the characteristics of each cluster and predict which observations belong to which clusters in order to choose the best number of clusters (*R* pkg. *randomForest*).
11. Keep the clusters that have a prediction error of less than .4 or more than 20 observations (985 out of 1050 observations) for the typology

3. Results

We begin by presenting the 7 dimensions (variable combinations) and then describe the typology they in turn provide.

3.1. Composite variables or dimensions of personal networks

1. *Nonkin Interaction*, indexed by how many nonkin respondents named as social companions and how many nonkin they named overall (highly correlated), measures the extent to which interviewees were engaged outside the family. Other measures, such as the frequency of entertaining people at home, outside the home, going to cultural or sports events, correlate with these measures, but including them in the composite would make it much noisier.
2. *Nonkin Support* is indexed by the number of nonkin with whom respondents reported discussing personal matters, the number of nonkin they relied on for advice, and the number of nonkin who could be asked for an emergency loan. A score on this dimension reflects the extent to which the respondent depended on nonkin for support. Importantly, it is distinct from the Nonkin Interaction dimension.
3. *Kin Social Involvement/Support* is indexed by the number of relatives with whom respondents reported spending social time, the number providing practical help, and the number providing emotional support. In the case of family, social activity and support went together.
4. *Kin Proximity* assesses the extent to which respondents' kin lived nearby, combining the number of immediate family members respondents estimated to live in the area, the percentage of actual named kin living within five minutes' drive, and the percentage of them living within an hour's drive.
5. *Activity-Based* measures the extent to which respondents' ties entailed hobbies or activities apart from work or family life. It sums the fraction of alters whom the respondent described as sharing the same leisure activities, a measure of how often the respondent reported seeing such people socially, and the number of nonkin with whom the respondent discussed such pastimes.
6. *Religion-Based* measures the extent to which associates were distinctively rooted in shared religious activities. It sums a measure of homophily based on religion with the respondents' reports of how often they attended services.
7. *Work-Based* measures the extent to which colleagues were distinctively important. It sums a measure of homophily based on doing the same type of work as well as a measure of how often respondents reportedly spent social time with coworkers.

3.2. Types of networks

From these 7 dimensions (composite variables) we generated 11 distinctive clusters of respondents. Table 1 orders them by size in the rows. To aid interpretation, we scaled all the dimensions to a 0–1 range by subtracting the minimum from each value and dividing by the range. The cells contain the average score on each dimension for each network type. Most of the variables are left skewed and cannot be transformed into percentiles or a normal distribution without losing some information. Many individuals had a value of zero on these measures, which is substantively meaningful. Rather than distorting these values by imposing a distribution, we show the mean, as well as the dimension's 1st and 3rd quartile values (the last three rows of the table) to provide a sense of where the cluster averages fit into the overall distribution. To assist in reading Table 1, we underlined cluster averages lower than the 1st quartile and **bolded** averages greater than the 3rd quartile. For

Table 1
Average scores on seven dimensions for eleven network types.

Type	*N* cases	Dimension: Nonkin interaction	Nonkin support	Kin involve/support	Kin prox	Activity based	Relig.-based	Work based
1. Career-and-Friends	239	**0.72**	**0.32**	0.21	0.19	0.28	0.21	0.31
2. Family-and-Community	201	0.47	0.07	**0.51**	0.52	0.23	0.27	0.24
3. Family-Only	154	0.16	0.02	0.31	0.54	0.04	0.23	0.07
4. Untethered	77	0.53	0.07	0.15	0.14	0.12	0.06	0.15
5. Energetic	65	**0.76**	0.12	0.30	0.52	**0.47**	0.10	0.32
6. Withdrawn	61	0.49	0.06	0.19	**0.62**	0.12	0.29	0.08
7. Home-and-Church	50	0.60	0.03	0.28	0.31	0.07	**0.52**	0.10
8. Semi-Isolated	47	0.36	0.08	0.29	0.15	0.09	0.12	0.01
9. Nonkin-as-Kin	32	0.62	**0.41**	0.15	**0.61**	0.25	0.35	0.04
10. Sociable	31	**0.72**	0.09	0.36	**0.66**	0.07	0.14	0.06
11. Just Activities	28	0.49	0.12	0.15	0.38	**0.57**	0.21	0.13
1st quartile	–	*.37*	*0*	*.18*	*.17*	0	*0*	*0*
mean	–	*.53*	*.14*	*.30*	*.40*	*.21*	*.23*	*.20*
3rd quartile	–	*.71*	*.21*	*.40*	*.61*	*.33*	*.44*	*.37*

Note: Values represent averages for clusters on each of the 7 dimensions (composite variables). Values above the 3rd quartile for each dimension are in **bold**, while values below the 1st quartile percentile are underlined.

variables in which more than 25% of observations had a zero, we underlined one or two values close to 0.

We next describe the seven major types that each include at least five percent of the sample. We sort them into three higher-level types: kin-dominant, nonkin-dominant, and small-network types.

Kin-dominant types:

2. *Family-and-Community* (20%). Networks in this type had the highest average on kin involvement (well above the 3rd quartile); that is, respondents relied upon kin for both for companionship and support. In addition, respondents were connected, at moderate levels, to others through work, church, and activities (hence the term *community*).

3. *Family-Only* (16%). This network type reflected an average level of involvement with kin, little engagement with nonkin (scoring below the 1st quartile on both nonkin support and nonkin interaction) and few commitments that might facilitate meeting nonkin (lowest average on knowing people via extracurricular activities and on socializing with people from work).

7. *Home-and-Church* (5%) networks included many people who shared the respondent's religion and respondents with these networks reported frequent church attendance (the highest average of any group). Notably, too, nonkin provided little support (average near zero), suggesting that respondents with these networks largely relied on family.

Nonkin-dominant types:

1. *Career-and-Friends* (24% of sample). Networks of this type included relatively distant kin and respondents reported relatively low involvement with kin. The networks instead included many nonkin, notably coworkers, who provided ego considerable support and sociability (high averages on the nonkin support, nonkin interaction, and work-based dimensions).

5. *Energetic* (7%). This network-type included many people whom the respondents knew through leisure activities and work (high on the activity-based and work-based dimensions). The networks were distinctively high on the number of nonkin but no less than average in kin involvement/support.

Small-network types:

4. *Untethered* (8%) networks were distinguished by distant relatives, families with whom respondents report little involvement/support (below 1st quartile on kin proximity and kin involvement/support). Although an average number of nonkin interaction partners were in these networks, those nonkin reportedly provided little support to ego. We label these networks "untethered" because the respondents who reported them were far from family and tended not to know people through church, work, or activities.

6. *Withdrawn* (6%) networks included many nearby relatives with whom the respondents had little to do (low kin involvement). Like the Untethered, these networks included a moderate number of nonkin interaction partners who reportedly provided ego very little support, but unlike the Untethered networks, the Withdrawn networks had many kin nearby but few exchanges with kin.

Description of the four smaller clusters that could be predicted with high accuracy can be found in the footnotes.[4]

We next present some traditional network descriptors associated with each type. We do so to illustrate that the each network type has a sensible correspondence to underlying basic network attributes. Table 2 presents the network size (net size *n*), number of confidants (confidants *n*),[5] number of sociability partners (social *n*), percentage of the network that is kin (avg. pct. kin), age-adjusted years having known nonkin (avg. years known), and network density (avg. density, calculated for a subsample of respondents' alters) for each of the network types presented above. Age-adjusted years reflect the average number of years the respondent has known a subsample of nonkin in their network minus the average number of years respondents their age have known nonkin. Thus, some clusters will have negative values on avg. years known because those respondents tend to have had shorter relationships with nonkin as compared with other people their age.

These characteristics reinforce the picture we sketched. For instance, the average percentage kin for each type corresponds with what one would expect—Career-and-Friends and Untethered are low, while Family-and-Community and Family-Only are high. The patterns also square with our picture of the clusters. The network size for Career-and-Friends is high, but the density is low and

[4] *Semi-Isolated* (4.5%) networks include distant relatives but ones who seemed to provide ego with essential support. *Nonkin-as-Kin* (3%) networks include nearby kin who were not involved with ego and nonkin who provided ego with support. The tenth cluster we label *Sociable* (3%) because the egos in these networks reported high levels of kin involvement and sociable nonkin ties, but low levels of activities, work, religion and nonkin support. These appeared to be networks with high kin and nonkin involvement independent of social contexts for them. Finally, the *Just Activities* (3%) networks include alters who shared the same activities and hobbies as ego, but otherwise contained few social ties, suggesting networks that are particularly "focused" around hobbies and activities.

[5] Confidants were identified though a question inquiring about with whom the respondent discusses "personal matters" such as those "about someone you are close to or something you are worried about."

Table 2
Selected characteristics of each network type.

Type	Net size *n* (med.)	Confidants *n* (med.)	Social *n* (med.)	Avg. pct. kin	Avg. years known	Avg. density
1. Career-and-Friends	22	3	8	27.3	0.24	0.37
2. Family-and-Community	20	3	8	55.9	0.08	0.50
3. Family-Only	10	1	3	74.3	−0.69	0.64
4. Untethered	14	1	5	30.0	−0.99	0.31
5. Energetic	22	2	10	30.0	1.27	0.47
6. Withdrawn	13	1	5	33.3	−0.29	0.52
7. Home-and-Church	18	2	7	40.0	1.03	0.47
8. Semi-Isolated	14	1	4	54.5	−2.86	0.37
9. Nonkin-as-Kin	17	3	6	26.4	−0.08	0.44
10. Sociable	26	2	10	34.6	3.06	0.57
11. Just Activities	15	1	4	28.2	1.01	0.46

average time known is moderate, suggesting that associates are relatively recent and have not had a chance to meet one another. When we compare groups with a median of only one confidant, such as Family-Only, Untethered, and Withdrawn, other interesting differences emerge. The Family-Only cluster has the fewest number of social alters, and trouble keeping in touch with nonkin (average time known), but a relatively high density (probably due to kin knowing one another). From our earlier analyses, we know that the Untethered and Withdrawn differ in their proximity to kin, but here we can also see that the Untethered have lower density networks and have known alters for even less time than the Withdrawn.

3.3. Summary of network descriptions

This is an inductive exercise and we caution against seeing too much precision or generalizability in the findings. Nonetheless, this exercise not only yields seven major types of network patterns, but also (1) suggests a basic typology of egocentric networks and (2) points to what seem to be underlying factors that sort respondents' networks into clusters.

The sorting factors that emerge include *accessibility*—notably egos' distance from kin, but also their immersion on other contexts such as church, work, and leisure pursuits; general *social activity level*—some people have (or at least, report having) more ties and social engagement than others; and finally perhaps a *lifestyle* dimension that identifies people who are kin-attracted or kin-averse.

In addition, we should note that we found a hierarchy of network attributes as we refined our analytical process. The split between kin and nonkin was present from the start; the most important original variable the RF algorithm identified was the *total number of nonkin*.[6] This suggests that the number of nonkin a person meets and stays in touch with determines or reflects many other aspects of his or her personal network.

We can infer other social styles of egos that shape networks. Network types reflected whether respondents or their families had moved away. They also captured whether families that remained physically proximate had managed to stay in touch. Kin-dominant networks also tended to have other characteristics. For instance, they were more likely to also be shaped by religious involvement. Meanwhile, only a small proportion of people who were estranged from nearby kin had managed to rebuild a supportive network of friends.

Nonkin-dominant networks were associated with a different set of characteristics. They were heavily shaped by work and activities. They also tended to reflect a kin/nonkin tradeoff. Only a few energetic people stayed in touch with distant kin while still sustaining strong ties to nonkin or, if they had not moved away from kin, met and befriended nonkin while remaining an active part of nearby family life.

As we refined our procedure, we consistently observed a small group of people who had "replaced" kin with nonkin (in Footnote 4, the Nonkin-as-Kin type representing 3% of the sample), saying that they could rely upon nonkin over kin for emergency loans, advice and judgment. We also saw, in numerous runs, the usefulness of work, religion, and activities each as a means of characterizing one or two clusters. Generally, one cluster of people who were highly engaged with nonkin also had moderate or high contact with kin and could count on kin for practical and emotional support.

At a high level of abstraction, there are consistencies between our findings and those summarized by Antonucci et al. (2010)-kin-based, nonkin-based, and small networks. However, some differences emerge. For example, unlike previous studies that applied clustering, we do not find that reports of feeling "close" are important nor do such measures cluster with other variables to modify the picture.[7] Instead, closeness reports appear not to correlate well with other network measures—thus leaving them excluded from use in the final typology. Second, although respondents differ in the size of their networks, there were few who maintained large, active kin and nonkin networks simultaneously. About three percent of networks (labeled "Sociable" in Table 1) entailed high levels of kin involvement/support and sociable nonkin ties.

3.4. Ego characteristics associated with network types

Who had which types of networks? The next issue we address is the correlates of being in a particular network cluster. Although some of the attributes of egos that we examine, like gender and age, can be considered causal, the causal direction of others, such as education, income, and residential mobility, is uncertain.[8] We simply pursue an effort to describe the socio-demographic profiles generally associated with each of the major types. In addition to socio-demographic characteristics, we include two interviewer ratings in our examination: how active the respondent's life appeared to be, and the respondent's attitude during the interview. Both of these ratings help to explain the particular network type associated with each respondent.

Overall, personal networks comprised primarily of nonkin (i.e., Career-and-Friends; Energetic) were more commonly reported by egos who were younger, more educated, more often employed, more often unmarried, childless, and new to the neighborhood. Personal networks dominated by kin (i.e., Family-and-Community; Family-Only; Home-and-Church) tended to be reported by those

[6] See Table A2 in Supplementary data for the top ten variables RFs identified.

[7] See Table A2 in Supplementary data for importance measures.

[8] For example, people who are highly committed to family may curtail their education (and thus their income) in order to stay nearby.

who were more religious, had children, were married, and were retired or homemakers.

Table 3 presents a more detailed picture. An explanation of the table is given below.

Note: Table 3 presents characteristics of egos who fell into each network type. The significance levels reflect the results of a correlation coefficient test indicating whether membership in the cluster was associated with a high or low value of the variable. (We used Pearson's r for **age** and **orgs_n**, and used Kendall's tau for all others.) The values are variable averages per cluster. Many represent a proportion for each cluster, such as the proportion of egos in the cluster that is female in the case of the first column. Other variables:

- **Partner** indicates the proportion of people in the cluster who are married or in a serious relationship.
- **Work_FT**: the proportion of people who are working full time.
- **College**: the proportion with a college degree or higher.
- **No_HS**: the proportion that did not complete high school or a GED.
- **Income_lower**: The proportion with a household or personal income below $40,000 (inflation adjusted to 2015).
- **Child_home**: the proportion with a younger child living at home
- **Child_away**: the proportion with a child (possibly adult) living away from home.
- **Christian**: the proportion of people who self-described as Catholic or Protestant.
- **Orgs_n**: the average of is a count variable of the number of organizations in which ego was active (e.g., civic, hobby, business).
- **New_NBHD**: proportion of respondents who had moved to the neighborhood in the previous two years.
- **Abroad < 16**: proportion of the respondents who grew up outside the U.S. (immigrating age 16 or later).
- **Int_active**: the average of an interviewer rating of whether the respondent appeared to lead an active life.
- **Int_neg_att.**: average of interviewers' assessment of the extent to which respondents displayed a negative attitude during the interview.

Kin-dominant types:

2. *Family-and-Community* (20%). Egos with such networks were much more likely than others to be married (84%), have children (50% had a child at home, 41% had a child living elsewhere), declare themselves Christian (83%), and to be financially stable (only 22% were "low income").

3. *Family-Only* (16%). Respondents reporting such networks were generally older ($m = 49$), not working (only 29% had a full time job), and poorer (52% were "low income"). Interviewers tended to describe these respondents as inactive and, in unreported analyses, we found that these respondents tended to describe themselves as unhappy.

7. *Home-and-Church* (5%). This type is distinctively one reported by women (78% of the egos were women), the middle-aged ($m = 46.7$), those not working (only 32% had a full-time job), and people who tended to be better off financially (only 16% were "low income").

Although these respondents and the types of networks they reported are similar, those with *Family-Only* networks tended to be disadvantaged, many being single parents, while the *Home-and-Church* respondents appeared to enjoy many advantages.

The *Family-and-Community* respondents appear to have been a younger, less established version of *Home-and-Church* ones, with both parents working to make ends meet.

Nonkin-dominant types:

1. *Career-and-Friends* (24% of sample). Given their high educational attainment, lack of children, above-average incomes, and the fact that they were often recently arrived (Table 3), it is likely that many of the people reporting such networks had moved to pursue a career and were trying to establish friendships and find a partner.

5. *Energetic* (7%). Respondents reporting such networks were younger (34 years), tended work full time (72%), and likelier to be male (57% versus a sample average of 44%). Interviewers said that they led very active lives.[9] Compared to *Career-and-Friends* respondents, these people were not as educated and not as mobile, but about as engaged with nonkin—while also managing to keep up with family and pursue leisure activities.

Small-network types:

4. *Untethered* (8%): Individuals with these networks were likelier to be male (57%), to be born abroad (14% compared to average of 7%), not religious,[10] and were more likely than people in any other cluster to be single, despite being middle-aged.

6. *Withdrawn* (6%). People with Withdrawn networks tended to be female (61%), also born abroad (20%), to be residentially stable (only 31% recently moved), and to be poor (39% were "low income"). Interviewers also reported that these respondents had the most negative attitudes during the interview.

These two groups are similar but those with *Withdrawn* networks were more likely to be poor, not working, less educated, and more negative during the interviews. They may have simply been a bit less personable than the *Untethered*, making it difficult for them to form supportive friendships or maintain strong relationships with the many kin who lived near them.

4. Discussion

We began with the question, What are the basic types of personal networks? In a large sample of northern Californians' ego networks in 1977, 11 basic types emerged. They are differentiated by egos' proximity to and involvement with kin, egos' social activity with and support from nonkin, and the extent to which egos knew people through work, church, or extracurricular activities. The two largest clusters are *Career-and-Friends* (24% of all ego networks) and *Family-and-Community* (20%). *Career-and-Friends* networks connected egos to distant kin, nonkin whom egos socialized with and also drew support from, many of whom they had met through work and extracurricular activities. Survey respondents with such networks tended to be relatively young, educated, single, and childless. *Family-and-Community* networks contained many nearby kin who were highly involved with egos, but also included alters connected to egos by work, church, and leisure activities. People reporting *Family-and-Community* networks tended to be middle-aged, married, parents, and avowed Christians.

To obtain these results, we relied heavily upon a method relatively new to sociology and social network analysis: Random Forests (RFs). RFs' ability to exploit a large number of combinations of variable values to find similarities and differences among observations made it more appealing than clustering methods traditionally used in the social sciences, such as latent class analysis. RFs offered the opportunity to capture a richer, more complex picture of personal networks. Using RF iteratively –to reduce many network descriptors into seven key dimensions and then using those to identify distinctive clusters—we could sort 94% of respondents into 11 meaningful network types whose members can be predicted with an average of 88% accuracy.

[9] We also found, in unreported analyses, that they considered themselves to be quite happy.

[10] Or simply not Christian – only 53% said they were Christian, but the network data in Table 1 indicate they have networks that are least involved with religion.

Table 3a
Ego attributes by network type (rotated).

Cluster	Female	Age	Partner	work_FT	College	no_HS	income_lower	child_home	child_away
1. Career-and-Friends	0.49*	37.7***	0.58***	0.65***	0.45***	0.03***	0.2***	0.23***	0.28***
2. Family-and-Comm.	0.62	39.3*	0.84***	0.57	0.19	0.12	0.22*	0.5***	0.41
3. Family-Only	0.6	49.0***	0.64	0.29***	0.06***	0.43***	0.52***	0.39	0.54**
4. Untethered	0.43*	44.2	0.57*	0.58	0.25	0.17	0.3	0.29	0.31
5. Energetic	0.43*	34.1***	0.69	0.72***	0.25	0.12	0.2	0.31	0.32
6. Withdrawn	0.61*	42.6	0.59	0.43	0.13*	0.15	0.39	0.34	0.34
7. Home-and-Church	0.78**	46.7*	0.7	0.32**	0.2	0.1	0.16*	0.44	0.46
8. Semi-Isolated	0.72*	52.4***	0.7	0.26	0.17	0.26	0.38	0.32	0.62***
9. Nonkin-as-Kin	0.66	35.9	0.63	0.44	0.31	0.06	0.34	0.5	0.16**
10. Sociable	0.48	40.7	0.84	0.58	0.32	0.06	0.29	0.39	0.29
11. Just Activities	0.36*	38.3	0.64	0.57	0.21	0.29	0.32	0.29	0.39
Mean	*0.56*	*41.6*	*0.67*	*0.52*	*0.25*	*0.16*	*0.29*	*0.36*	*0.38*

* <.05.
** <.01.
*** <.001.

Table 3b
Ego attributes by network type (cont.) (rotated).

Cluster	Christian	Orgs_n	New_NBHD	Abroad < 16	Int_active	Int_neg_att.
1. Career-and-Friends	0.55***	2.41***	0.54***	0.07	1.42***	1.43**
2. Family-and-Comm.	0.83***	2.12*	0.39	0.04*	1.25	1.47
3. Family-Only	0.84***	0.65***	0.35*	0.07	0.68***	1.82***
4. Untethered	0.53**	1.68	0.39	0.14*	1.08	1.69
5. Energetic	0.71	2.22	0.54	0.02	1.45***	1.48
6. Withdrawn	0.64	1.46	0.31	0.20***	1.18	1.85**
7. Home-and-Church	0.78	1.52	0.34	0.12	1.16	1.4
8. Semi-Isolated	0.6	1.15*	0.34	0.06	0.74***	1.32*
9. Nonkin-as-Kin	0.63	1.63	0.41	0.03	1.25	1.69
10. Sociable	0.71	1.84	0.45	0.03	1.26	1.23*
11. Just Activities	0.68	2.11	0.36	0.04	0.96	1.61
Mean	*0.69*	*1.79*	*0.42*	*0.07*	*1.16*	*1.55*

* <.05.
** <.01.
*** <.001.

While we believe this represents an advance in constructing typologies of personal networks, some caveats are in order. The use of inductive clustering, rather than a deductive approach such as regression or classification, reflects our belief there are internal logics particular to the data that are worth discovering. A problem with such inductive work is that the emergent typologies from two different datasets will not be completely commensurable. They may have somewhat different measures. Moreover, some measures will mean different things in one population than in another. Thus, we can only compare the results of studies based on clustering after we have substantively interpreted them. In a sense, this is an "interpretive sociology."

Moreover, clustering procedures are sensitive to changes in the data and variables because they depend upon the strength of similarities within groups and dissimilarities between groups. Had we discovered and added other variables or had the sampling frame been different, our results might have been somewhat different. Despite those caveats, we believe that a comparable approach to constructing a typology with a comparable sample and features would still point to critical differentiators similar to those here: accessibility of kin, companionship and support from kin vs. nonkin, and the contexts for seeing people, such as work, church, and leisure activities. Other factors, such as exchange multiplexity or the duration of ties may also emerge as new differentiators.

5. Conclusion

Our findings extend and refine the typologies described in previous studies. Antonucci et al. (2010) conclude that most network studies find a diverse and superficial network type, a socially-isolated type, a friend or community type, and a family type. Previous studies relied upon self-assessed "closeness" or frequency of interaction to determine which relationships were superficial versus intimate. In contrast, we did not find self-assessed closeness to be a useful measure; it did not align with other features of respondents' personal networks in our data. We instead relied upon recent interactions and exchanges to measure socializing and support.

Other differences between our findings and previous findings can be seen as extensions. We found four types that resembled Antonucci et al.'s friend/community clusters. Our largest type, *Career-and-Friends* (24%) best matches a "diverse" (and diffuse) network type that Antonucci et al. note tends to be the largest in previous studies. Yet we also find an *Energetic* type (7%), networks of respondents who engaged with many people in many contexts. We found a third type, the *Sociable* (3%), networks of both kin and nonkin sociability that did not rest upon contexts such as work, church, or leisure activities. And, lastly, we found a network type, *Nonkin-as-Kin* (3%), in which nonkin had replaced the supportive and sociability functions of family.

A second set of extensions relates to family-centric clusters. The largest family type (*Family-and-Community*, 20%) departs from previous findings in that its networks combine a high representation of family with many nonkin known through leisure activities, work, and church. The *Family-Only* (16%) type accords more closely with the inward-looking family cluster in some other studies. Egos in these networks knew few nonkin and counted upon almost none of them for support. Interestingly, *Family-Only* networks contained fewer relatives than the more active *Family-and-Community* did, perhaps because egos with the former networks were poorer and more stressed and thus less able to sustain ties even to family. *Home-and-Church* (5%) networks include nonkin who were known socially but not relied upon, strong family ties, and strong church connections.

Lastly, we expanded upon the socially-isolated type mentioned by Antonucci et al. We found three types that were somewhat isolated: one with few kin even though kin lived nearby (*Withdrawn*, 6%); one with few ties, but including kin even though the kin were geographically distant (*Semi-Isolated*, 5%); and one lacking family ties, perhaps because of geographic mobility, and including nonkin–but only for sociability, not support (*Untethered*, 8%).

In addition to building upon previous studies, our study serves as a reminder of how closely bound up seemingly distinct features of an individual's social experience are. For instance, it suggests that distance from kin and commitments to work, religious organizations, and extracurricular hobbies condition the opportunities for connecting to kin and nonkin.

The key to weaving together these seemingly distinct features is a better grasp of networks in a life-course perspective. We believe that network typologies might mesh well with a life-course perspective, such as the convoy model (Antonucci and Akiyama, 1987; Antonucci et al., 2014). For instance, we mentioned geographic and cultural factors that make certain network types more likely than others. People who stay near family can get caught up in caring for both young children and aging parents; those who move for the sake of their careers tend to end up with companions and supporters from work and other activities, perhaps while delaying marriage and children. Combining a typological approach with something like the convoy model may allow researchers to identify common stories of network formation.

The issue that we could not explore in depth, which is equally important, is the role of cultural norms and expectations in shaping what exchanges take place in these different contexts and, as a result, what relationships people are able to form and maintain (see Fiori et al., 2008 for such an example among the elderly in the U.S. and Japan). While straightforward comparisons between populations may be possible using a pre-constructed classification scheme and fitting individuals to it, or by simply comparing individual variable measures, we think an inductive approach as applied in this study may better preserve cultural and circumstantial differences. Identifying, as we have done here, the bundles of network characteristics unique to different populations would allow for comparison of the circumstantial and cultural factors that shape personal networks.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.socnet.2016.02.003.

References

Antonucci, T.C., Akiyama, H., 1987. Social networks in adult life and a preliminary examination of the convoy model. J. Gerontol. 42 (5), 519–527.

Antonucci, T.C., Ajrouch, K.J., Birditt, K.S., 2014. The convoy model: explaining social relations from a multidisciplinary perspective. Gerontologist 54 (1), 82–92.

Antonucci, T.C., Fiori, K.L., Birditt, K., JackeyF L.M.H., 2010. Convoys of social relations: integrating life-span and life-course perspectives. In: The Handbook of Life-Span Development. John Wiley & Sons Inc.

Armstrong, J.S., 1967. Derivation of theory by means of factor analysis or Tom Swift and his electric factor analysis machine. Am. Stat. 21 (5), 17–21.

Bailey, S., Marsden, P.V., 1999. Interpretation and interview context: examining the general social survey name generator using cognitive methods. Soc. Netw. 21, 287–309.

Barnes, J.A., 1954. Class and committees in a Norwegian Island Parish. Hum. Relat. 7 (1), 39–58.

Bastani, S., 2007. Family comes first: men's and women's personal networks in Tehran. Soc. Netw. 29 (3), 357–374.

Bearman, P., Parigi, P., 2004. Headless frogs and other important matters: conversation topics and network structure. Soc. Forces 83, 535–557.

Birditt, K.S., Antonucci, T.C., 2007. Relationship quality profiles and well-being among married adults. J. Family Psychol. 21 (4), 595–604.

Bott, E., 1955. Urban families: conjugal roles and social networks. Hum. Relat. 8, 345–384.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Breiman, L., 2003. Manual – Setting Up, Using, And Understanding Random Forests V4.0, Available from: https://www.stat.berkeley.edu/~breiman/Using_random_forests_v4.0.pdf [accessed 10.08.15].

Blum, T., 1985. Structural constraints on interpersonal relations: a macrosociological analysis. Am. J. Sociol. 91 (3), 511–521.

Caruana, R., Niculescu-Mizil, A.,2006. An empirical comparison of supervised learning algorithms. In: Proceedings of the 23rd International Conference on Machine Learning, ICML'06. ACM, New York, NY, USA, pp. 161–168.

Chavent, M., Kuentz, V., Liquet, B., Saracco, J., 2012. ClustOfVar: an R package for the clustering of variables. J. Stat. Softw. 50 (13), 1–16.

Coleman, J., Katz, E., Herbert, M., 1957. The diffusion of an innovation among physicians. Sociometry 20 (4), 253–270.

Cutler, A., Cutler, D.R., Stevens, J.R., 2012. Random forests. In: Zhang, C., Ma, Y. (Eds.), Ensemble Machine Learning. Springer, US, pp. 157–175 http://link.springer.com/chapter/10.1007/978-1-4419-9326-7_5.

Díaz-Uriarte, R., de Andrés, S.A., 2006. Gene selection and classification of microarray data using random forest. BMC Bioinform. 7 (1), 3.

Feld, S., 1984. The structured use of personal associates. Soc. Forces 62, 640–652.

Fiori, K.L., Antonucci, T.C., Cortina, K.S., 2006. Social network typologies and mental health among older adults. J. Gerontol. Ser. B: Psychol. Sci. Soc. Sci. 61 (1), P25–P32.

Fiori, K.L., Antonucci, T.C., Akiyama, H., 2008. Profiles of social relations among older adults: a cross-cultural approach. Ageing Soc. 28, 203–231.

Fischer, C.S., 1982. To Dwell Among Friends: Personal Networks in Town and City. University of Chicago Press.

Genuer, R., Poggi, J.-M., Tuleau-Malot, C., 2010. Variable selection using random forests. Pattern Recog. Lett. 31 (14), 2225–2236.

Gibney, S., McGovern, M.E., 2011. Social support networks and mental health: evidence from share. J. Epidemiol. Commun. Health 65, 1–43.

Granovetter, M.S., 1973. The strength of weak ties. Am. J. Sociol. 78 (6), 1360–1380.

Grimmer, J., King, G., 2011. General purpose computer-assisted clustering and conceptualization. Proc. Natl. Acad. Sci. 108 (7), 2643–2650.

Hennig, M., 2007. Re-evaluating the community question from a German perspective. Soc. Netw. 29 (3), 375–390.

Jain, A.K., 2010. Data clustering: 50 years beyond k-means. Pattern Recog. Lett. 31 (8), 651–666.

Kim, J., 2012. The configurations of social network members and social support in old age. In: Paper presented to the Population Association of America, San Francisco, CA.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 23 (3).

Marks, S.R., 1994. Intimacy in the public realm: the case of co-workers. Soc. Forces 72, 843–858.

Merton, R.K., 1968. Social Theory and Social Structure. Simon and Schuster.

Pescosolido, B.A., Wright, E.R., Alegría, M., Mildred, V., 1998. Social networks and patterns of use among the poor with mental health problems in Puerto Rico. Med. Care 36 (7), 1057–1072.

Rook, K., 1987. Social support versus companionship: effects on life stress, loneliness, and evaluation by others. J. Personal. Soc. Psychol. 52, 1132–1147.

Shi, T., Seligson, D., Belldegrun, A.S., Palotie, A., Horvath, S., 2005. Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma. Modern Pathol. 18 (4), 547–557.

Small, M.L., 2013. Weak ties and the core discussion network: why people regularly discuss important matters with unimportant alters. Soc. Netw. 35, 470–483.

Smith, J.A., McPherson, M., Smith-Lovin, L., 2014. Social distance in the united states sex, race, religion, age, and education homophily among confidants, 1985 to 2004. Am. Sociol. Rev. 79 (3), 432–456.

Wellman, B., 1979. The community question: the intimate networks of east yorkers. Am. J. Sociol. 84 (5), 1201–1231.

Wellman, B., Gulia, M., 1999. The network basis of social support: a network is more than the sum of its ties. In: Barry Wellman, B. (Ed.), Networks in a Global Village. Westview Press, Colo, pp. 83–118.

Zachary, W.W., 1977. An information flow model for conflict and fission in small groups. J. Anthropol. Res. 33 (4), 452–473.