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into types: career-and-friends (24%), family-and-community (20%), family-only (16%), untethered (8%), energetic (7%), withdrawn (6%), and home-and-church (5%).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.