ELSEVIER

Social Networks 24 (2002) 407–422

SOCIAL NETWORKS

www.elsevier.com/locate/socnet

Egocentric and sociocentric measures of network centrality

Peter V. Marsden

Department of Sociology, Harvard University, 630 William James Hall, 33 Kirkland Street, Cambridge, MA 02138, USA

Abstract

Egocentric centrality measures (for data on a node's first-order zone) parallel to Freeman's [Social Networks 1 (1979) 215] centrality measures for complete (sociocentric) network data are considered. Degree-based centrality is in principle identical for egocentric and sociocentric network data. A closeness measure is uninformative for egocentric data, since all geodesic distances from ego to other nodes in the first-order zone are 1 by definition. The extent to which egocentric and sociocentric versions of Freeman's betweenness centrality measure correspond is explored empirically. Across seventeen diverse networks, that correspondence is found to be relatively close—though variations in egocentric network composition do lead to some notable differences in egocentric and sociocentric betweennness. The findings suggest that research design has a relatively modest impact on assessing the relative betweenness of nodes, and that a betweenness measure based on egocentric network data could be a reliable substitute for Freeman's betweenness measure when it is not practical to collect complete network data. However, differences in the research methods used in sociocentric and egocentric studies could lead to additional differences in the respective betweenness centrality measures.

Keywords: Centrality; Egocentric networks; Betweenness

1. Introduction

Because it resonates so strongly with conceptions of social power and structural influence, centrality is among the properties of social networks most often invoked in applied studies. To judge by the subsequent scholarly attention given to them, Freeman's late-1970s articles on measuring centrality stand as the most prominent of his many contributions to understanding social networks. As of November 2001, the 'Social Sciences Citation Index' (Institute for Scientific Information, 2001) records 289 citations to "Centrality in social networks. I. Conceptual clarification" (Freeman, 1979) in the sources it tracks, and an additional 83 citations to "A set of measures of centrality based on betweenness" (Freeman,

E-mail address: pvm@wjh.harvard.edu (P.V. Marsden).

PII: S0378-8733(02)00016-3

1977). Beyond these classic papers explicating centrality measures for binary network data, Freeman's influential work on centrality includes experimental studies (Freeman et al., 1980) and generalizations to valued data (Freeman et al., 1991). Others have developed the conceptions of centrality in Freeman's work in many different directions (Faust and Wasserman, 1992; Faust, 1997).

Freeman's (1979) measures based on degree, betweenness, and closeness are defined for "complete" or "sociocentric" network data that provide information on relationships among all nodes within a bounded social network. This design yields the rich data on indirect connectedness among nodes required by two of the measures, and is generally regarded as the preferred design for network data. Because it can be difficult to implement the sociocentric design for a large network, however, many network studies instead use an "egocentric" design that obtains information about only that portion of a network in the immediate locality of a given node. The egocentric design does not require a priori enumeration of a population of nodes, and is often used to measure social networks in survey-based studies.

This article examines egocentric versions of Freeman's sociocentric centrality measures. It first considers special cases of these measures for egocentric data on a node's first-order zone. In principle, the sociocentric and egocentric designs yield identical information on degree centrality. However, the closeness centrality measure is uninformative when based on the egocentric design. The remainder of the article compares the egocentric and sociocentric versions of Freeman's betweenness measure. When these measures are calculated based on the same network data, differences between them yield insight into the information about centrality lost through using the egocentric design. Their correspondence is studied here for seventeen networks from diverse empirical settings. In most of these, egocentric betweenness is very closely associated with Freeman's sociocentric measure. We examine some conditions under which the two correspond less well, and briefly compare egocentric betweenness to some other measures based on egocentric data. The analyses suggest that—apart from possible differences due to the distinct data collection methods typically used in egocentric and sociocentric studies—egocentric betweenness often may be a reliable substitute for Freeman's sociocentric betweenness measure.

2. Centrality measures for egocentric data

This article considers only the case of binary symmetric data on a single relation. The sociocentric network data assumed, therefore, consist of a single binary who-to-whom matrix A of dimension $N \times N$, where N is the number of nodes, with element $a(p_i, p_k) = a(p_k, p_i) = 1$ if nodes p_i and p_k are related to one another, and 0 otherwise; diagonal elements $a(p_i, p_i)$ are 0 by convention. The egocentric networks of interest include those nodes and relationships within the first-order zone surrounding a given "ego" node p_i. If the nodes of A constitute the population that can appear within node p_i's egocentric network, the corresponding egocentric data matrix A_i is an N_i by N_i submatrix of A, where N_i is the size (including ego) of the egocentric network. A_i consists of p_i, nodes in the first-order zone of p_i (i.e. nodes p_k for which $a(p_i, p_k) = 1$), elements $a(p_i, p_k)$ linking node p_i to nodes in its first-order zone, and elements $a(p_j, p_k)$ (where $a(p_i, p_k) = a(p_i, p_j) = 1$) relating nodes in p_i's first-order zone.[1]

With these definitions in mind, we review Freeman's (1979) centrality measures and their special cases for egocentric data. The egocentric version of each measure is obtained by calculating it for node p_i (the ego node) within the egocentric data matrix A_i.

2.1. *Degree centrality*

"Degree" centrality is measured simply as the number of direct ties that involve a given node. Freeman (1979, p. 221) gives a conceptual interpretation for this measure: degree reflects the direct relational activity of node p_i. For sociocentric data:

$$C_{\mathrm{D}}(p_i) = \sum_{k=1}^{N} a(p_i, p_k)$$

Since (by definition) the egocentric network A_i for node p_i includes all other nodes p_k such that $a(p_i, p_k) = 1$, degree centrality measures based on egocentric and sociocentric data are in principle identical. Both give network size for node p_i, $N_i - 1$.

2.2. *Closeness centrality*

Freeman's closeness centrality measures depend on the geodesic distance $d(p_i, p_k)$, that is, the minimal length of an indirect path from p_i to p_k. Conceptually, Freeman (1979, p. 224) suggests that measures of centrality based on closeness reflect a node's freedom from the controlling actions of others, their capacity for independent action within the network. A "farness" measure based on geodesic distances for sociocentric data is

$$C_{\mathrm{C}}(p_i)^{-1} = \sum_{k=1}^{N} d(p_i, p_k)$$

Farness is defined only for connected components of a graph, i.e. subsets of nodes that are mutually connected at some finite geodesic distance. Small values of $C_{\mathrm{C}}(p_i)^{-1}$ identify highly central nodes according to this criterion. A corresponding closeness measure is based on the reciprocal of $C_{\mathrm{C}}(p_i)^{-1}$:

$$C'_{\mathrm{C}}(p_i) = \frac{N-1}{\sum_{k=1}^{N} d(p_i, p_k)}$$

Since the egocentric network A_i for node p_i includes another node p_k if and only if $a(p_i, p_k) = 1$, all geodesic distances $d(p_i, p_k)$ from the ego node p_i to other nodes p_k within the egocentric network are necessarily 1. $C_{\mathrm{C}}(p_i)^{-1}$, therefore, equals $N_i - 1$ for egocentric data, while $C'_{\mathrm{C}}(p_i)$ equals 1. Therefore, with the egocentric design, a centrality measure based on closeness provides no information beyond that given by a measure based on degree.

[1] Not all instruments for collecting egocentric data elicit data on relationships among elements in the first-order zone, but such information is essential for defining egocentric betweenness.

2.3. Betweenness centrality

Centrality measures based on betweenness reflect the intermediary location of a node along indirect relationships linking other nodes. Freeman (1979, p. 221) gives a conceptual interpretation for betweenness measures in coordination/control terms: a node with high betweenness has a capacity to facilitate or limit interaction between the nodes it links. Freeman's betweenness measure for sociocentric data is

$$C_{\mathrm{B}}(p_i) = \sum_{j=1}^{N}\sum_{k=1}^{j-1} b_{jk}(p_i)$$

where $b_{jk}(p_i)$ (defined as $g_{jk}(p_i)/g_{jk}$, where g_{jk} is the total number of geodesic paths linking p_j and p_k, and $g_{jk}(p_i)$ is the number of those geodesic paths that include p_i) reflects the extent to which node p_i lies between nodes p_j and p_k.

A measure of betweenness $C_{\mathrm{B}}(p_i)$ based on egocentric network A_i will correspond imperfectly to the sociocentric version.[2] Usually, egocentric betweenness will be smaller than sociocentric betweenness. With network data from an egocentric design, the maximal geodesic distance between points p_j and p_k is 2, since by definition all pairs of nodes within the first-order zone of p_i are connected either directly or indirectly via p_i (as well as, possibly, via other "alter" nodes within p_i's egocentric network). The egocentric betweenness measure will not reflect p_i's intermediary location between indirectly tied pairs of nodes (some of which lie outside of node p_i's first-order zone) linked by geodesic distances of length 3 or more. The latter indirect relationships are, however, counted by the sociocentric betweenness measure.

In another way, however, a node's betweenness can be misleadingly high when assessed using the egocentric design. When $a(p_j, p_k) = 0$, the egocentric betweenness measure will exaggerate the extent to which p_i lies between p_j and p_k if there exist one or more nodes p_m not linked directly to p_i (and, therefore, not within egocentric network A_i) for which $a(p_m, p_j) = a(p_m, p_k) = 1$. In the sociocentric design, such nodes p_m are also intermediaries for the p_j–p_k relationship, but geodesics involving such nodes are omitted from the count of geodesics g_{jk} when a betweenness measure is based on egocentric data. Such omissions bias $b_{jk}(p_i)$ upwards. We conjecture, however, that this upward bias in egocentric betweenness due to omission of nodes like p_m from egocentric networks usually will be smaller than the downward bias due to the omission of many geodesic paths of length longer than 2.

[2] The egocentric measure of betweenness for p_i is $C_{\mathrm{B}}(p_i)$ calculated using data on egocentric network A_i rather than sociocentric network A. A simpler egocentric betweenness measure is the number of pairs of nodes (not involving ego) in the egocentric network that are not directly connected to each other (and, therefore, are indirectly connected via a geodesic through ego); this simpler measure is, in essence, the reverse of egocentric network density:

$$\sum_{j\in A_i, j\neq i}^{N_i} \sum_{k\in A_i, k\neq i}^{j-1} [1 - a(p_j, p_k)]$$

The egocentric analog of Freeman's betweenness measure additionally takes account of the fact that some pairs (p_j, p_k) may be connected not only via the ego node, but also through other nodes in the egocentric network that thereby share the intermediary position for that relationship with the ego node.

There is strong reason to anticipate that the correspondence of egocentric and sociocentric betweenness measures, while imperfect, is nonetheless substantial. First, the measures have a part–whole correlation, since geodesics of length 2 that lie within egocentric network A_i are a subset of all geodesic paths on which node p_i lies. Moreover, longer geodesic paths involving p_i that do not lie within A_i nonetheless contain length-2 geodesics within A_i; a node that lies on many length-2 geodesics within the egocentric network will tend to lie on many longer geodesics involving nodes outside of A_i.

The extent to which egocentric and sociocentric measures of betweenness centrality are correlated certainly varies from one network to another. If the two measures generally correspond closely, however, then egocentric betweenness may be a serviceable substitute for the sociocentric version when the data required to calculate the latter are not available. In Section 3, we investigate this issue empirically using several familiar sets of network data.

3. How closely do egocentric and sociocentric betweenness correspond?

This section examines the impact of study design on betweenness centrality by comparing egocentric and sociocentric betweenness measures calculated for the same network data. We begin with sociocentric (complete) network data, and calculate Freeman's measure as a reflection of the "true" betweenness of nodes. Next, successively for each node, we treat the same data as if they had been assembled using an egocentric design and calculate egocentric betweenness. We then examine the correlation of the sociocentric and egocentric measures to ascertain the extent to which information about the relative betweenness of nodes is lost through using the egocentric design rather than the sociocentric one. These comparisons assume that study design is the sole reason for differences between the sociocentric and egocentric versions of the data.[3]

All of the networks studied here are binary and symmetric.[4] Sociocentric centrality measures were calculated for the complete network data using Ucinet 5 (Borgatti et al., 1999); egocentric centrality for each node p_i was calculated for the subnetwork A_i obtained by extracting p_i and its first-order zone from the complete network.[5]

3.1. Example: egocentric centrality for the Bank Wiring Room

The classic Roethlisberger and Dickson (1939, p. 501) data set on game-playing relationships in the Bank Wiring Room is the *Dropsophilia* fruit fly for social network studies, though it is more often used to illustrate the identification of subgroups than the relative centrality of nodes. The familiar data are displayed in Fig. 1.[6] The network includes two near-cliques linked by a single bridging tie from W5 to W7, and two isolates (S2 and I3).

---

[3] This need not be so, since research methods for assembling egocentric and sociocentric data often differ. This issue is discussed further in Section 5.

[4] Initially nonsymmetric data were made symmetric taking the p_j–p_k relationship to be present if either $a(p_j, p_k) = 1$ or $a(p_k, p_j) = 1$.

[5] The egocentric measure was calculated using a stand-alone FORTRAN program available from the author upon request. It is possible to calculate the egocentric measure via Ucinet 5, but it is cumbersome to do so since this requires the extraction of a submatrix containing each node and its first-order zone separately for each ego.

[6] Plots in Figs. 1 and 3 were produced via KrackPlot 3.0 (Krackhardt et al., 1994).

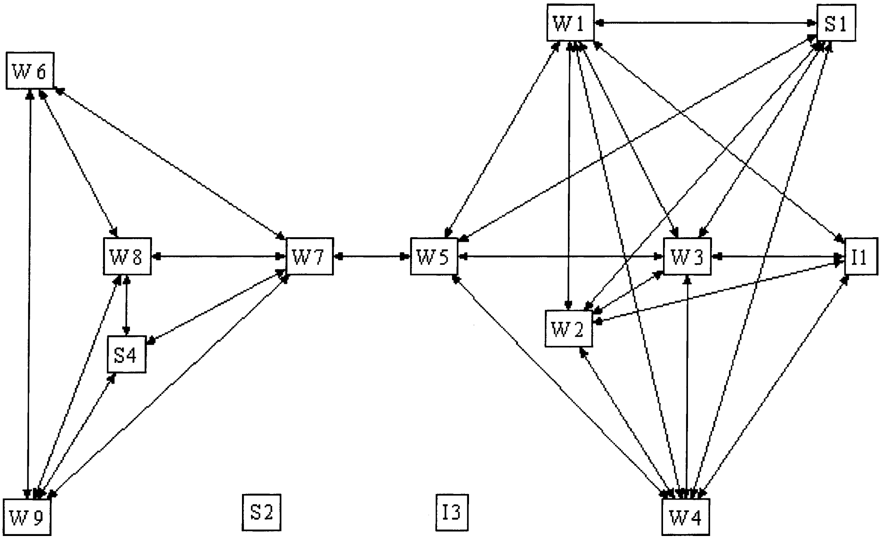

Fig. 1. The Bank Wiring Room–games network.

Because the W5–W7 tie is the only bridge joining the cliques, the actors at either end of it are highly central within the network, lying on all between-clique geodesics.

Table 1 presents egocentric and sociocentric betweenness centrality for each node in the Bank Wiring Room. Consider W7, the actor with the highest egocentric betweenness score of 4.33. W7 serves as the gatekeeper (Freeman, 1980) to the smaller clique, lying on all geodesic paths to the larger one. W7's egocentric betweenness reflects his unique intermediary location along four two-step paths (to W5, from W6, W8, W9, and S4), together with shared betweenness (with W8 and W9) for the W6–S4 relationship.

Table 1
Egocentric and sociocentric betweenness for the Bank Wiring Room–games network

Node	Sociocentric betweenness	Egocentric betweenness	Effective size
W1	3.75	0.83	2.00
W2	0.25	0.25	1.40
W3	3.75	0.83	2.00
W4	3.75	0.83	2.00
W5	30.00	4.00	2.60
W6	0.00	0.00	1.00
W7	28.33	4.33	3.00
W8	0.33	0.33	1.50
W9	0.33	0.33	1.50
S1	1.50	0.25	1.40
S2	0.00	0.00	0.00
S4	0.00	0.00	1.00
I1	0.00	0.00	1.00
I3	0.00	0.00	0.00

Because egocentric betweenness is based on geodesic paths of length 2 only, values of the egocentric betweenness measure are generally smaller than their sociocentric counterparts. Node W5, for example, has betweenness centrality 30 on the sociocentric measure, but only 4 on the egocentric measure. Notwithstanding this difference in scale, however, the measures based on the two study designs rank nodes in the network similarly in terms of their relative centrality. Nodes W5 and W7 are most central by both measures. Nodes with 0 sociocentric betweenness (W6, S2, S4, I1, and I3) also have 0 egocentric betweenness. Even the relative centrality of sets of nodes having intermediate values ((W1, W3, W4) and (W2, W8, W9, S1)) is similar for the two measures. Across the 14 nodes, the Pearson correlation of egocentric and sociocentric betweenness is 0.993 (see Table 2). Thus, for this clearly structured network, egocentric betweenness reflects the comparative centrality of nodes very well.

3.2. Comparisons for larger networks

The Bank Wiring Room "games" network has a particularly clear structure, which makes it ideal for illustrating network measures. Because the structure in most other networks is not as self-evident, one must exercise caution before generalizing findings for this particular network to others.

The network in Fig. 1 has at least two properties that would seem to make likely a close correlation between egocentric and sociocentric betweenness: relatively small size and comparatively high network density. By limiting the length of geodesic paths, these features make it likely that length-2 geodesic paths (on which the egocentric measure depends)

Table 2
Correspondence of egocentric betweenness with sociocentric betweenness, degree, and effective size

Network	Number of nodes	Correlation of egocentric betweenness with:		
		Sociocentric betweenness	Degree	Effective size
Bank Wiring Room–games	14	0.993	0.377	0.790
Krackhardt managers–advice	21	0.977	0.851	0.933
Krackhardt managers–friendship	21	0.995	0.910	0.962
Cottage 4 women–admiration	25	0.913	0.910	0.960
Rail Town physicians–multiplex	35	0.874	0.777	0.941
River Town physicians–multiplex	44	0.827	0.876	0.947
Old Town physicians–multiplex	50	0.993	0.918	0.945
Altneustadt elites–social	51	0.919	0.900	0.940
Altneustadt elites–business	51	0.883	0.879	0.912
Altneustadt elites–civic	51	0.965	0.885	0.918
Lazega lawyers–friendship	71	0.959	0.790	0.907
Lazega lawyers–advice	71	0.982	0.868	0.951
Lazega lawyers–coworker	71	0.978	0.892	0.955
Big Town physicians–multiplex	117	0.975	0.935	0.944
Twin Cities charities–director acquaintance	196	0.973	0.877	0.923
Twin Cities charities–resource flow	198	0.941	0.862	0.891
Webster dormitory–friendship	217	0.973	0.873	0.928

constitute a relatively high fraction of geodesic paths of any length (on which the sociocentric measure rests). This will in turn raise the part–whole correlation between the two measures.

For this reason, the comparison of betweenness measures shown in Table 1 was made for networks drawn from a diverse set of other studies. The studies were selected in order to ensure considerable variation in network size. They include the advice and friendship networks linking 21 managers in a high-technology organization (Krackhardt, 1987; Wasserman and Faust, 1994, pp. 740–741), a sociometric network of admiration ties among 25 women in the New York State Training School for Girls (Forsyth and Katz, 1946), networks linking physicians in four small cities in the midwestern US (Coleman et al., 1966; Burt, 1987),[7] the business-professional, social relations, and civic discussion networks among 51 community leaders in the small German city of Altneustadt (Laumann and Pappi, 1976), the advice, friendship, and coworker networks for 71 attorneys in a New England law firm (Lazega, 1999), networks of resource exchange and inter-director acquaintance among public charities in the Minneapolis–St. Paul area (Galaskiewicz and Bielefeld, 1998), and a network of citations as "friends" or "best friends" for 217 students in a dormitory at an Australian university (Webster, 1995). Table 2 presents results of comparing egocentric and sociocentric betweenness measures for 17 networks from these studies.

Correlations of egocentric and sociocentric betweenness scores reported in Table 2 are generally quite high, usually above 0.9. They do not decline markedly with network size. In fact, the lowest egocentric–sociocentric correlations in Table 2 are for physician networks from the two smaller cities in the 'Medical Innovation' study ($r = 0.827$ for the "River Town" network, $N = 44$; $r = 0.874$ for the "Rail Town" network, $N = 35$) and the business-professional network in the Altneustadt elite study ($r = 0.883$, $N = 51$). Egocentric and sociocentric measures of betweenness are correlated at or above 0.94 for the four networks of size over 100; for the Webster dormitory data ($N = 217$), the measures share nearly 95% of their variance.[8]

With the exception of the network involving the 44 River Town physicians, a node's egocentric betweenness is more closely associated with its sociocentric betweenness than with its degree, as we see by comparing the second and third columns of Table 2. The generally close correspondence of the two betweenness measures, then, reflects more than the fact that both are correlated with network size.

3.3. *What conditions produce a weak correspondence between egocentric and sociocentric betweenness?*

It is clear that egocentric and sociocentric betweenness will be identical in some kinds of networks. In a complete (fully connected) graph, for example, all nodes actors will have

[7] Analyses of the four physician networks are based on a multiplex network constructed as the union of advice, discussion, and friendship linkages.

[8] It is possible that treatments of the data other than that used here (see Footnote 4) might lead to a lower egocentric–sociocentric correspondence than shown in Table 2. Use of the "maximum" rule for symmetrizing the data raises a network's density and lowers its diameter, thus making length-2 geodesics a larger fraction of all geodesics. The overall densities of the symmetric networks studied here do vary quite considerably, however, from 0.021 (Twin Cities resource flows) to 0.691 (Krackhardt managers–advice), and variations in density are not obviously associated with the size of the egocentric–sociocentric correlation.

egocentric and sociocentric betweenness 0, since there are no geodesics longer than 1 in such a graph. In a maximally centralized (star) graph (Freeman, 1979, p. 219), egocentric and sociocentric betweenness will likewise be identical: the egocentric network for the central node in such a graph is identical to the sociocentric network, while the egocentric networks for peripheral nodes (having 0 betweenness) consist of a single tie to the central node.

The correlations displayed in Table 2 indicate that egocentric and sociocentric betweenness are not identical in any of the networks studied here. Nonetheless, they show that sociocentric betweenness can be approximated more or less well as a linear function of egocentric betweenness. There are, however, nodes for which egocentric and sociocentric betweenness correspond poorly—that is, for which predicted betweenness based on the first-order zone substantially over- or understates betweenness within the complete network.

To explore some conditions under which sociocentric betweenness is disproportionately higher or lower than predicted based on the egocentric measure, we examined scatterplots for the networks having comparatively low egocentric–sociocentric correlations (Table 2), focusing on cases with large residuals. One such plot—for the network of business-professional ties among community leaders in Altneustadt—is displayed in Fig. 2. This illustrates the clear positive relationship between the two betweenness measures. Some nodes, however, lie relatively far from the regression line.

The largest residuals in Fig. 2 are for nodes 1 (Berghaus) and 27 (K. Koenig). Koenig lies well above the regression line, while Berghaus is located substantially beneath it. Koenig has greater sociocentric betweenness than anticipated on the basis of his egocentric betweenness, while Berghaus's egocentric betweenness converts into less sociocentric betweenness than

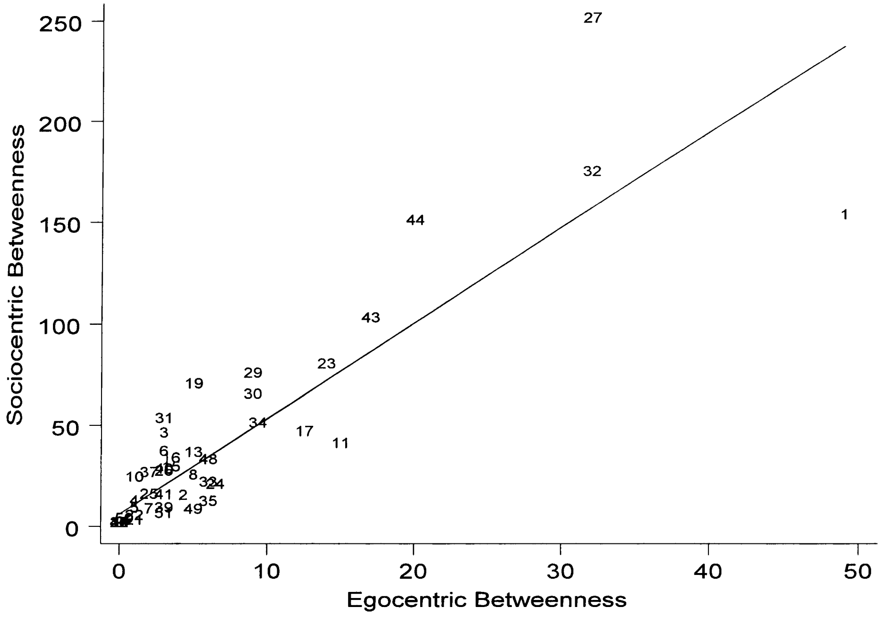

Fig. 2. Scatterplot of sociocentric vs. egocentric betweenness, for Altneustadt business-professional network (numbers in plot identify nodes).

typical within this network. These two egocentric networks were examined closely and are shown in Fig. 3.

Berghaus has the largest egocentric network of any node in this network (11 alters), and the highest egocentric betweenness (at 49). The regression line for this network (sociocentric =

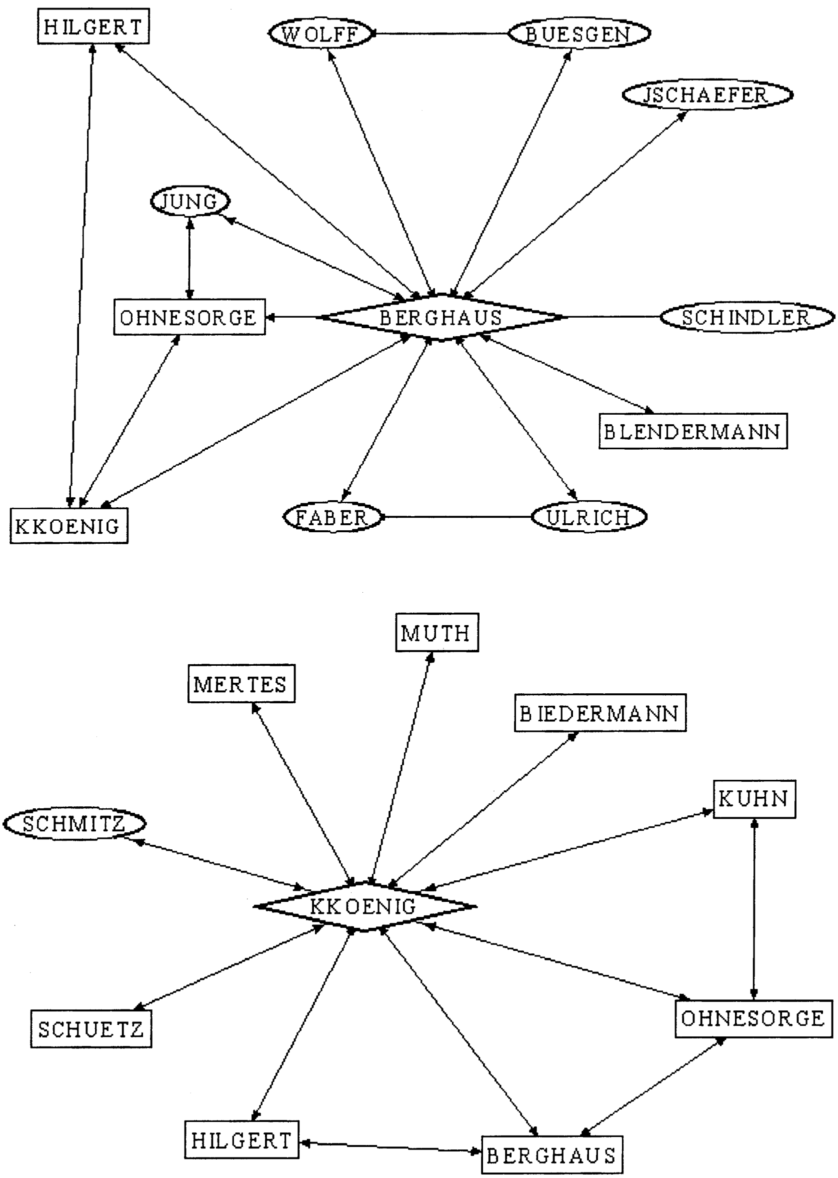

Fig. 3. Egocentric networks for nodes 1 (Berghaus) and 27 (K. Koenig) in Altneustadt business-professional network. Diamonds identify ego nodes. Rectangles identify alter nodes with above-average sociocentric betweenness. Ellipses identify alter nodes with below-average egocentric betweenness.

$5.9 + 4.7 \times$ egocentric) predicts that Berghaus would have sociocentric betweenness 237, but in fact Berghaus lies along only 152 geodesics in the full network. Koenig, on the other hand, has a smaller (nine alter) egocentric network and egocentric betweenness 32. This would convert into sociocentric betweenness 157 according to the fitted regression, but Koenig actually lies on nearly 250 geodesic paths according to Freeman's betweenness measure calculated using the full network data.

The key feature that differentiates the egocentric networks of Berghaus and Koenig is the centrality of the alters within their first-order zones. In the two egocentric networks plotted in Fig. 3, alters are distinguished according to whether they are above average (rectangles) or below average (ellipses) in sociocentric betweenness. Koenig's egocentric network is composed almost exclusively of high-centrality alters, while most of Berghaus's more numerous contacts are low-centrality alters. Moreover, many of Berghaus's high-centrality contacts are interconnected and, therefore, somewhat redundant, while most of Koenig's high-centrality alters are not directly linked to one another. This means that the two-step geodesics on which Koenig lies will be leveraged into more intermediary positions on longer geodesics than will those on which Berghaus lies, producing the greater increase of the sociocentric measure over the egocentric one seen for Koenig in Fig. 2.

In their analysis of eigenvector centrality measures, Mizruchi et al. (1986) drew a distinction between "hubs" and "bridges." This distinction captures a difference between the two egocentric networks in Fig. 3, and suggests one set of conditions under which we should observe relatively large divergences between egocentric and sociocentric betweenness. Mizruchi et al. (1986, pp. 30–31) note that a node "can become central in one of two ways: it can interlock with many peripheral units (we call it a *hub*), or it can interlock with a few central units (we call it a *bridge*)." Nodes having high hub centrality and low bridge centrality will have egocentric networks resembling that of Berghaus, and lower sociocentric betweenness than would be expected on the basis of their egocentric betweenness. Nodes having high bridge centrality (like Koenig), on the other hand, will have higher values on Freeman's betweenness measure than anticipated on the basis of their egocentric betweenness.[9] Similar differences in the composition of egocentric networks were also evident for nodes having large residuals in plots for other networks (the River Town and Rail Town physicians) for which Table 2 reports relatively low egocentric–sociocentric correlations.

4. Measures for egocentric networks related to egocentric betweenness

Several measures closely related to egocentric betweenness have been proposed previously in the network literature. This section briefly discusses several of them.

Most directly related is the measure of brokerage presented by Gould and Fernandez (1989, pp. 96–102). They define brokerage in terms of a node's location along a two-step path linking two other nodes. Most of Gould and Fernandez's discussion focuses on specific types of brokerage which also depend on labels (subgroup membership) assigned to

[9] Koenig's egocentric network exhibits both hub and bridge form. Its comparatively large size gives it hub form, but it is the bridge form due to its inclusion of high-centrality alters that accounts for Koenig's disproportionately high sociocentric betweenness.

nodes in the network under study. However, their "total brokerage score" (p. 102), which aggregates across five specific forms of brokerage, is equivalent to egocentric betweenness as defined previously. Gould and Fernandez calculate their measures using sociocentric network data (see Fernandez and Gould, 1994) and do not report comparisons of egocentric and sociocentric betweenness.

Conceptually related to egocentric betweenness is Burt's (1992, p. 52) index giving the "effective size" of a node's egocentric network. Burt gives this measure for node i as

$$\sum_{j}^{N_i} \left[1 - \sum_{q\ (q \neq i,j)}^{N_i} x_{iq} m_{jq} \right]$$

where x_{iq} reflects node p_i's network time and energy invested in a relationship with p_q while m_{jq} indicates the strength of the relationship between p_j and p_q. A contact with p_j is redundant for p_i to the extent that the summation inside the brackets is large; redundant contacts reduce the effective size of an egocentric network. Strong relationships m_{jq} between alter nodes p_j and p_q within p_i's egocentric network thus lower its effective size. Such relationships also reduce p_i's egocentric betweenness.

Egocentric betweenness and effective size are not identical, since effective size is based on continuous measures of investment and tie strength, while egocentric betweenness is defined for binary network data. The two measurements are nonetheless closely associated, as we see in Table 2. In all networks examined aside from the Bank Wiring Room, egocentric betweenness and effective size are correlated at or above 0.89.[10] Egocentric betweenness is more closely associated with effective size than with degree in all 17 networks.[11]

Fischer (1982, p. 291) also developed a measure of centrality for egocentric network data. This measure, however, referred to the relative centrality of the alters within an egocentric network, rather than the comparative centrality of different egos. Fischer's measure for alters is "the number of other network ... members each one knew well."

5. Conclusion and caveats

This article has considered versions of Freeman's (1979) centrality measures for egocentric network data on a node's first-order zone. Degree-based measures are in principle identical for egocentric and complete (sociocentric) network data. Closeness measures are uninformative for egocentric data, since all geodesic distances from ego to other nodes in the first-order zone are 1 by definition. Most of the article has explored empirically the extent to which egocentric and sociocentric versions of Freeman's betweenness centrality measure correspond. Across seventeen diverse networks varying considerably in size, that correspondence has been found to be relatively close—though variations in egocentric

[10] In calculating effective size, binary data on egocentric networks were transformed to measures x_{iq} and m_{jq} as suggested by Burt (1992, p. 51) and as implemented in Ucinet 5 (Borgatti et al., 1999).

[11] The comparatively weak association for the Bank Wiring Room appears to be due to a nonlinear relationship produced by the fact that the two isolates have effective size 0. The correlation of effective size and egocentric betweenness is 0.91 for the 12 nonisolate nodes in this data set.

network composition do lead to some notable divergences of egocentric betweenness from its sociocentric counterpart.

The implication is that a betweenness measure based on egocentric network data could be a reliable substitute for Freeman's betweenness measure when it is not practical to collect complete network data. It is, of course, possible that the correspondence of egocentric and sociocentric betweenness might be weaker in networks that are much larger than those examined here. There was not, however, a markedly weaker egocentric–sociocentric correlation in the larger networks studied above.

The comparisons presented have examined the effect of study design on centrality under the assumption that the egocentric and complete designs produce the same data on egocentric networks. This assumption is questionable for several reasons, and these concluding comments remark on some of them. To the extent that the egocentric data obtained under the two designs differ, measures of degree will differ, and the egocentric and sociocentric betweenness measures will diverge more than Table 2 suggests.

The most common survey methods for gathering sociocentric network data (see Marsden, 1990) ask respondents to report on their outgoing ties using a recognition format that reminds respondents of all alters deemed to be part of the network under study. By contrast, typical survey methods for assembling egocentric network data enumerate alters via free recall, and rely on the respondent (the ego node) as an informant about the structure and composition of the entire first-order zone (see Burt, 1984).

How might these differences in research methods affect the egocentric betweenness measure? Consider first the use of recall versus recognition for eliciting elements of the first-order zone. Research on forgetting (Brewer, 2000; Brewer and Webster, 1999) shows that survey respondents do not recall a substantial fraction of the contacts they subsequently recognize. This means that data from the usual egocentric design typically understate degree, and—other things being equal—will understate egocentric betweenness by omitting some geodesics of length 2.

Next, consider differences that might arise by virtue of the fact that egos usually provide all data on the first-order zone in the egocentric design, while the sociocentric network design measures connections among alters using reports from the alters themselves. Freeman (1992) suggests that informants tend to perceive social affiliations in categorical terms; they impose transitivity when reporting on social ties by "filling in the blanks." Applied to egocentric data, this suggests that informants will tend to overstate the strength of ties linking alters in their first-order zones (see Kumbasar et al., 1994, p. 497), and thereby understate their egocentric betweenness by reporting that alters are linked directly rather than indirectly via the ego.

The conclusion that egocentric betweenness will be lower when measured using the usual research methods for assembling egocentric network data than with those typical for sociocentric data seems at variance with the findings of Kumbasar et al. (1994) on biases in social perception. Using cognitive social structure (CSS) data (Krackhardt, 1987), Kumbasar et al. present a persuasive case that individuals tend to perceive themselves as being more central within a group than they are regarded as being by others in that group. In particular, individuals tend to claim to have more contacts than others perceive them to have.

Kumbasar et al. (1994) identify important regularities in social perception, but their findings do not pertain directly to understanding differences between research methods typical of egocentric and sociocentric studies. The comparison of interest here is that between a node's egocentric betweenness in the first-order zone of that node's self-reported "slice" (Krackhardt, 1987, p. 115) of CSS data (the data usually obtained for egocentric networks), and the same node's egocentric betweenness in its first-order zone for data from what Krackhardt (1987, p. 116) terms a locally aggregated structure, "the type of data normally collected in traditional sociometry." For both of these forms of network data, a node is the informant providing survey information on its own direct ties.[12] Measures based on both are, therefore, affected in the same way by a tendency for nodes to report more ties for themselves than others see them as having. Given this, other perceptual biases reported by Kumbasar et al. (1994)—for informants to overstate network density and transitivity within their first-order zones—would have the effect of reducing a node's egocentric betweenness.

At least one practical consideration would complicate efforts to assess betweenness on the basis of egocentric network data, notwithstanding effects of perceptual biases. Measurement of large egocentric networks using the ego node as an informant pose appreciable respondent burden. Respondents are asked to provide name interpreter data on relationships for many pairs of alters—a task that is not only time-consuming but tedious. Name generator-based approaches to measuring egocentric networks (Burt, 1984) often manage this respondent burden by obtaining name interpreter data on no more than a fixed number of alters, but that would lead to a serious understatement of egocentric betweenness for those nodes with a number of alters exceeding the threshold.

In closing, we raise one conceptual question about betweenness centrality. Implicit in our use of Freeman's sociocentric betweenness measure as a standard for assessing the usefulness of the egocentric betweenness measure is an assumption that Freeman's measure reflects a node's "true" centrality. This in turn means that an intermediary position on a short geodesic is treated as being equivalent to a similar location on a longer geodesic. In many social networks, however, there must be coordination among the intermediary nodes along geodesics of length greater than 2 if they are to profit from their positions (Marsden, 1982). There hence exist barriers to the use of any control potential flowing from such a location that do not apply to length-2 geodesics. If nodes can make effective use only of their control over the short indirect paths on which they lie (see Gould and Fernandez, 1989), then egocentric betweenness provides a good measure of this local control potential within a network, rather than an approximation to a "true" betweenness centrality captured by Freeman's sociocentric measure.

Acknowledgements

For helpful comments, I am grateful to Devon Brewer and an anonymous reviewer. For access to some of the network data examined above, I am indebted to Ronald S. Burt, Joseph Galaskiewicz, Edward O. Laumann, Emmanuel Lazega, and Cynthia Webster.

[12] The Kumbasar et al. (1994) comparisons in essence contrast individual "slices" of a CSS with what Krackhardt (1987, p. 117) calls "consensus structures" that define each dyadic tie on the basis of perceptions of all respondents. Moreover, the data in the Kumbasar et al. (1994) study were obtained using recognition methods, and hence do not reflect the phenomenon of forgetting.

References

Borgatti, S.P., Everett, M.G., Freeman, L.C., 1999. Ucinet 5 for Windows: Software for Social Network Analysis. Analytic Technologies, Harvard, MA.

Brewer, D.D., 2000. Forgetting in the recall-based elicitation of personal and social networks. Social Networks 22, 29–43.

Brewer, D.D., Webster, C.M., 1999. Forgetting of friends and its effects on measuring friendship networks. Social Networks 21, 361–373.

Burt, R.S., 1984. Network items and the General Social Survey. Social Networks 6, 293–339.

Burt, R.S., 1987. Social contagion and innovation: cohesion versus structural equivalence. American Journal of Sociology 92, 1287–1335.

Burt, R.S., 1992. Structural Holes: The Social Structure of Competition. Harvard University Press, Cambridge, MA.

Coleman, J.S., Katz, E., Menzel, H., 1966. Medical Innovation: A Diffusion Study. Bobbs-Merrill, Indianapolis.

Faust, K., 1997. Centrality in affiliation networks. Social Networks 19, 157–191.

Faust, K., Wasserman, S., 1992. Centrality and prestige: a review and synthesis. Journal of Quantitative Anthropology 4, 23–78.

Fernandez, R.M., Gould, R.V., 1994. A dilemma of state power: brokerage and influence in the national health policy domain. American Journal of Sociology 99, 1455–1491.

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

Forsyth, E., Katz, L., 1946. A matrix approach to the analysis of sociometric data: preliminary report. Sociometry 9, 340–347.

Freeman, L.C., 1977. A set of measures of centrality based on betweenness. Sociometry 40, 35–41.

Freeman, L.C., 1979. Centrality in social networks: I. Conceptual clarification. Social Networks 1, 215–239.

Freeman, L.C., 1980. The gatekeeper, pair-dependency and structural centrality. Quality and Quantity 14, 585–592.

Freeman, L.C., 1992. Filling in the blanks: a theory of cognitive categories and the structure of social affiliation. Social Psychology Quarterly 55, 118–127.

Freeman, L.C., Roeder, D., Mulholland, R.R., 1980. Centrality in social networks. II. Experimental results. Social Networks 2, 119–141.

Freeman, L.C., Borgatti, S.P., White, D.R., 1991. Centrality in valued graphs: a measure of betweenness based on network flow. Social Networks 13, 141–154.

Galaskiewicz, J., Bielefeld, W., 1998. Nonprofit Organizations in an Age of Uncertainty: A Study of Organizational Change. Aldine de Gruyter, New York.

Gould, R.V., Fernandez, R.M., 1989. Structures of mediation: a formal approach to brokerage in transaction networks. In: Clogg, C.C. (Ed.), Sociological Methodology 1989. Basil Blackwell Ltd., Cambridge, MA, pp. 89–126.

Institute for Scientific Information, 2001. Social Sciences Citation Index. Institute for Scientific Information, Philadelphia.

Krackhardt, D., 1987. Cognitive social structures. Social Networks 9, 109–134.

Krackhardt, D., Blythe, J., McGrath, C., 1994. KrackPlot 3.0: an improved network drawing program. Connections 17 (2), 53–55.

Kumbasar, E., Romney, A.K., Batchelder, W.H., 1994. Systematic biases in social perception. American Journal of Sociology 100, 477–505.

Laumann, E.O., Pappi, F.U., 1976. Networks of Collective Action: A Perspective on Community Influence Systems. Academic Press, New York.

Lazega, E., 1999. Generalized exchange and economic performance: social embeddedness of labor contracts in a corporate law partnership. In: Leenders, R.T.A.J., Gabbay, S.M. (Eds.), Corporate Social Capital and Liability. Kluwer, Boston, pp. 237–265.

Marsden, P.V., 1982. Brokerage behavior in restricted exchange networks. In: Marsden, P.V., Lin, N., (Eds.), Social Structure and Network Analysis. Sage Publications, Beverly Hills, pp. 201–218.

Marsden, P.V., 1990. Network data and measurement. Annual Review of Sociology 16, 435–463.

Mizruchi, M.S., Mariolis, P., Schwartz, M., Mintz, B., 1986. Techniques for disaggregating centrality scores in social networks. In: Tuma, N.B. (Ed.), Sociological Methodology 1986. American Sociological Association, Washington, DC, pp. 26–48.

Roethlisberger, F.J., Dickson, W.J., 1939. Management and the Worker. Harvard University Press, Cambridge, MA.
Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press, New York.
Webster, C., 1995. Detecting context-based constraints in social perception. Journal of Quantitative Anthropology 5, 285–303.