On the robustness of centrality measures under conditions of imperfect data

Stephen P. Borgatti [a,*], Kathleen M. Carley [b,1], David Krackhardt [c,1]

[a] *Carroll School of Management, Boston College, Department of Organization Studies, Fulton Hall, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA*
[b] *Institute for Software Research International, Carnegie Mellon University, USA*
[c] *Heinz School of Public Policy, Carnegie Mellon University, USA*

Abstract

An analysis is conducted on the robustness of measures of centrality in the face of random error in the network data. We use random networks of varying sizes and densities and subject them (separately) to four kinds of random error in varying amounts. The types of error are edge deletion, node deletion, edge addition, and node addition. The results show that the accuracy of centrality measures declines smoothly and predictably with the amount of error. This suggests that, for random networks and random error, we shall be able to construct confidence intervals around centrality scores. In addition, centrality measures were highly similar in their response to error. Dense networks were the most robust in the face of all kinds of error except edge deletion. For edge deletion, sparse networks were more accurately measured.
© 2005 Published by Elsevier B.V.

1. Introduction

Conventional wisdom has it that network analysis is intolerant of missing or mistaken data. On the basis of examples like Fig. 1 (Roethlisberger and Dickson, 1939), we typically reason that the failure to record a single node (such as W5) or a single tie (such as the tie between W5 and W7) can lead to a radically different understanding of the network and

---

* Corresponding author. Tel.: +1 617 552 0450; fax: +1 617 552 4230.
  E-mail addresses: borgatts@bc.edu (S.P. Borgatti), kathleen.carley@cmu.edu (K.M. Carley).
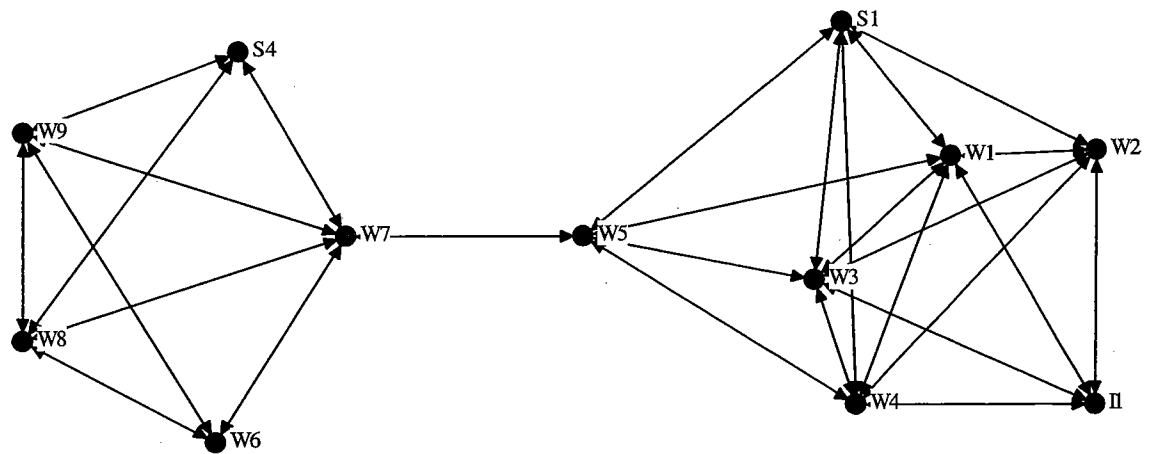 [1] Tel.: +1 412 268 6016; fax: +1 412 268 1744.

Fig. 1. Games relation from Hawthorne Bank Wiring Room study (Roethlisberger and Dickson, 1939). Isolates removed.

misleading measurements of network indices such as centrality. Similarly, it is commonly assumed that it is not possible to measure node properties like centrality accurately when sampling nodes or ties from networks.

However, very little work has actually been done to check these assumptions and evaluate under what circumstances measures computed on "errorful" or sampled network data could be accurate. A notable exception is the work by Marsden (1990) using the GSS data on networks. In later work, Marsden (1993) went on to suggest conditions under which network density and composition estimates were likely to be reliable. More recently, Costenbader and Valente (2003) examined the robustness of centrality measures in the face of inaccurate or incomplete network data. In this paper, we replicate and extend this work, examining the accuracy of centrality measures when data are incomplete due to either random measurement error (including missing data) or deliberate sampling on nodes or edges.

Understanding the robustness of basic network measures is extremely important in order to assess the validity of network research. This is particularly true when we study large or covert networks, where the data is likely to be missing or hidden (Carley et al., 2001; Carley, 2003). Centrality measures provide a good starting point as they are among the most frequently used of all network measures. Our fundamental research question is to understand how accuracy in the measurement of centrality declines with increasing error. As a practical corollary, we ask how the relationship between error and robustness is affected by basic characteristics of the network, such as size and density.

We should note that previous work in this area (Marsden, 1990; Costenbader and Valente, 2003) relied on empirical samples (i.e., "real data") in assessing the various network measures. A limitation of this approach is that the sampling errors contained in the data are likely to be systematic, but the pattern is unknown. Another limitation is that the sample of networks is necessarily very limited. To overcome these limitations, we take a statistical computational approach and examine robustness in a very large sample of random graphs, into which we introduce a controlled amount and kind of error.

2. Methodology

We start with a known or true network (which could also be regarded as the population network), compute centrality (referred to as "true centrality"), distort or sample from the network to generate the observed network, compute centrality on the observed network (called the "observed centrality"), and then compare the true and observed centrality measures.

To construct the true networks, we used random networks of varying sizes and densities generated by the method of Erdos and Renyi (1959). Networks of size 10, 25, 50, and 100 nodes were generated. For each size, networks of the following densities were constructed: 1, 2, 5, 10, 30, 50, 70, and 90%.

For each network, exactly one of four types of error was introduced in order to construct the observed network: node removal, node addition, edge removal, and edge addition. Node removal refers to the extraction of a certain proportion of existing nodes, selected at random. Node addition refers to the insertion of a certain proportion of extra nodes into the network. Whenever a node was added, ties were also randomly added from the node to other nodes in the network. The degree of the new node was chosen by randomly selecting an existing node and adopting its degree.[1] Edge removal refers to dropping a given proportion of edges, selected at random. Edge addition refers to the addition of ties not present in the true network, simulating measurement error in which ties are erroneously recorded between nodes that in truth are not tied. The following proportions of error were introduced across all four kinds of error: 0, 1, 5, 10, 25, and 50%.[2]

Both the node-removal and edge-removal cases can be thought of as forms of sampling from the network, since what remains after removal is a random sample of the network. However, we note that this kind of sampling is different from the kind of sampling found in ego-network analysis, in which all ties of a given type connected to a sampled node are collected, including ties to nodes not in the sample. For measuring degree centrality and possibly betweenness centrality (Everett and Borgatti, 2005), this ego-net form of sampling would be expected to yield more accurate centrality estimates than the kind of sampling we are doing here.

Thus, for each kind of error, we use an $8 \times 6 \times 4$ factorial design (eight levels of density × six levels of error × four levels of network size) for a total of 192 distinct combinations. For each of the 192 combinations, we generated 10,000 pairs of true and observed networks.

There are many measures of centrality of a node in a graph (Wasserman and Faust, 1994, p. 169), dating back as far as Moreno (1934) who identified "stars" as those with special importance and influence. Within this family, four measures stand out as foundational in our field: degree, closeness, betweenness, and eigenvector centralities. Their prominence within the field of network analysis stems from the fact that they all have strong yet distinct theoretical underpinnings (Freeman, 1979; Bonacich, 1972, 1987) and that they are frequently used for empirical analysis of social systems (e.g., see Brass and Burkhardt (1993) for a review of their use in organizational research; see Podolny (1993) for an influential study

---

[1] However, ties among the new nodes add to each node's degree, so that new nodes have higher overall degree than corresponding original nodes.

[2] The 0% error cases provided a check on the programming and a starting point for line diagrams.

Table 1
Measures of centrality robustness

Measure	Description
Top 1	Proportion of times that the most central node in the true network is also the most central node in the observed network
Top 3	Proportion of times that the most central node in the observed network is among the top three most central nodes in the true network
Top 10%	Proportion of times that the most central node in the observed network is among the top ten percent of nodes in the true network
Overlap	Number of nodes in both the top 10% of the true network and the top 10% of the observed network, divided by the number of nodes in either
R^2	Square of the Pearson correlation between true centralities and observed centralities, taking as cases only nodes found in both the true and observed networks

based on Bonacich's eigenvector centrality). For these reasons, we will restrict ourselves to these four measures in the present paper.

For each pair of networks, we measure degree, betweenness, closeness and eigenvector centrality, and compare them using five measures of accuracy (see Table 1). The accuracy measures are averaged across the 10,000 replications for each cell in the experimental design, resulting in a single value for each measure for each of the 192 factorial combinations.

The first measure of accuracy, labeled "Top 1" in tables all tables, is the average proportion of times that the most central node in the observed network was also the most central node in the true network. The second measure, labeled "Top 3", is the proportion of trials in which the most central node in the observed network is among the top three most central nodes in the true network. The third measure, labeled "Top 10%", is the proportion of trials in which the most central node in the observed network is in the top decile of the true network. The fourth, labeled "Overlap", gives the overlap between the top decile of the observed network and the top decile of the true network. It is computed as $|U \cap V|/|U \cup V|$, where U is the set of nodes in the top decile of the observed network, and V is the set of nodes in the top decile of the true network. The numerator gives the size of the intersection of U and V, while the denominator gives the size of the union of U and V. The ratio is 1 when the deciles are identical, and 0 when they are wholly disjoint. The last measure of accuracy, labeled "R^2", is the square of the Pearson correlation coefficient between true and observed centrality. We interpret it as the proportion of variance in true scores accounted for by observed centrality. For the node-removal and node-addition error types (in which true and observed networks have different sets of nodes), the correlation measure is calculated on only those nodes that appear in both the true and observed networks; nodes not in common are ignored.

The analysis consists primarily of scatterplots of centrality robustness as a function of amount of measurement error added, across different types of error. In addition, some regressions are performed in order to examine the effects of network size and density on robustness. The data for these regressions are mean accuracy scores for each of the 192 experimental combinations; hence, the cases are treatment groups. As a result, the regressions are descriptive only, and we do not report significance tests.

3. Results

Table 2 presents average accuracy scores for all centrality measures and all measures of accuracy at all error levels of all types of error. For simplicity, only results for graphs of size 100 and density 50% are shown.[3] The effects of network size and density on robustness are discussed in a later section.

Perhaps the most striking thing about Table 2 is the similarity in results obtained across centrality measures. A comparison of corresponding columns for each measure (such as node removal for degree compared to node removal for betweenness) shows that the four centrality measures behave virtually identically in the face of measurement error. This suggests that the distinction between local and global measures of centrality (Scott, 2000) is not as important as previously thought. These results are consistent with those of Everett and Borgatti (2005) who found that betweenness calculated on ego networks (a local measure) was, on average, nearly identical to betweenness calculated on the full network in which ego networks were embedded (a global measure).

Fig. 2 presents scatter plots of betweenness accuracy as a function of error level for all four kinds of error. Each curve in the plots represents a different method of measuring accuracy. We note that the accuracy measures have distinctive characteristic curves. The Top 1 and Overlap measures decline the most rapidly, in a pattern resembling an exponential decay. These are the strictest measures of accuracy. The R^2 and Top 3 measures decline approximately linearly with increasing error. Finally, the Top 10 measure of accuracy is also largely linear and the least sensitive, declining quite gently as a function of error.

It is also apparent that accuracy declines very smoothly and predictably with increases in error. For example, the R^2 measure declines linearly with error, and in fact the proportion of variance in true centrality accounted for by observed centrality is approximately 1 − Error, so that losing 25% of nodes yields 75% variance accounted for, while losing 50% of nodes yields 50% variance accounted for. The monotonicity and predictability of accuracy decline is very good news because it means that if we knew the level of (random) error in a dataset, we could in principle construct error bounds around our centrality estimates.

The effects of different kinds of error are remarkably similar to each other. In particular, the shapes of the accuracy functions are nearly the same across error types. With respect to overall levels of accuracy, careful comparison of the values in Table 2 shows that node errors are generally more forgiving than edge errors at the same nominal error rate, and that node addition is in general the most forgiving error while edge addition is the least forgiving. A regression analysis (not shown) of accuracy on (dummied) error type, controlling for error level, network size and density, shows that on average, node addition increases the expected R^2 by approximately .25 relative to edge addition.

An intriguing area to explore is the effect of density on accuracy. Table 3 shows standardized regression coefficients for density in regressions of betweenness accuracy on density and error level, for graphs of size 100.[4] As noted earlier, the cases for the regressions are the

---

[3] Because we investigate both the addition and deletion of ties, 50% density represents a neutral point for comparison with respect to robustness. As explained in a later section, lowering density favors edge removal, while raising it favors edge addition.

[4] Results are similar for the other measures of centrality.

Table 2
Accuracy results for graphs of 100 nodes and 50% density

% Error	Degree centrality				Betweenness				Closeness				Eigenvector			
	Node removal	Node addition	Edge removal	Edge addition	Node removal	Node addition	Edge removal	Edge addition	Node removal	Node addition	Edge removal	Edge addition	Node removal	Node addition	Edge removal	Edge addition
Top 1																
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	.86	.88	.83	.84	.86	.86	.82	.82	.86	.88	.83	.84	.86	.87	.83	.83
5	.68	.71	.62	.62	.68	.69	.61	.62	.68	.71	.62	.62	.68	.71	.61	.61
10	.57	.62	.49	.49	.56	.58	.48	.48	.57	.62	.49	.49	.56	.62	.49	.48
25	.34	.47	.29	.30	.34	.41	.28	.29	.34	.47	.29	.30	.34	.49	.29	.29
50	.16	.35	.15	.14	.15	.27	.14	.13	.16	.35	.15	.14	.16	.38	.15	.14
Top 3																
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1.00	1.00	.99	.97	.99	.99	.99	.97	1.00	1.00	.99	.97	.99	1.00	.99	.97
5	.94	.95	.91	.87	.93	.94	.90	.86	.94	.95	.91	.87	.93	.95	.91	.87
10	.86	.89	.80	.77	.85	.87	.80	.75	.86	.89	.80	.77	.85	.89	.80	.76
25	.63	.75	.55	.53	.63	.70	.54	.51	.63	.75	.55	.53	.62	.78	.54	.53
50	.34	.61	.32	.29	.33	.51	.30	.28	.34	.61	.32	.29	.34	.65	.31	.29
Top 10%																
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	1.00	1.00	1.00	.99	1.00	1.00	1.00	.98	1.00	1.00	1.00	.99	1.00	1.00	1.00	.99
10	.99	1.00	.98	.95	.99	.99	.98	.95	.99	1.00	.98	.95	.99	1.00	.98	.95
25	.91	.96	.85	.81	.90	.94	.85	.79	.91	.96	.85	.81	.91	.97	.85	.80
50	.64	.88	.61	.55	.63	.81	.59	.54	.64	.88	.61	.55	.64	.91	.59	.54

Table 2 (*Continued*)

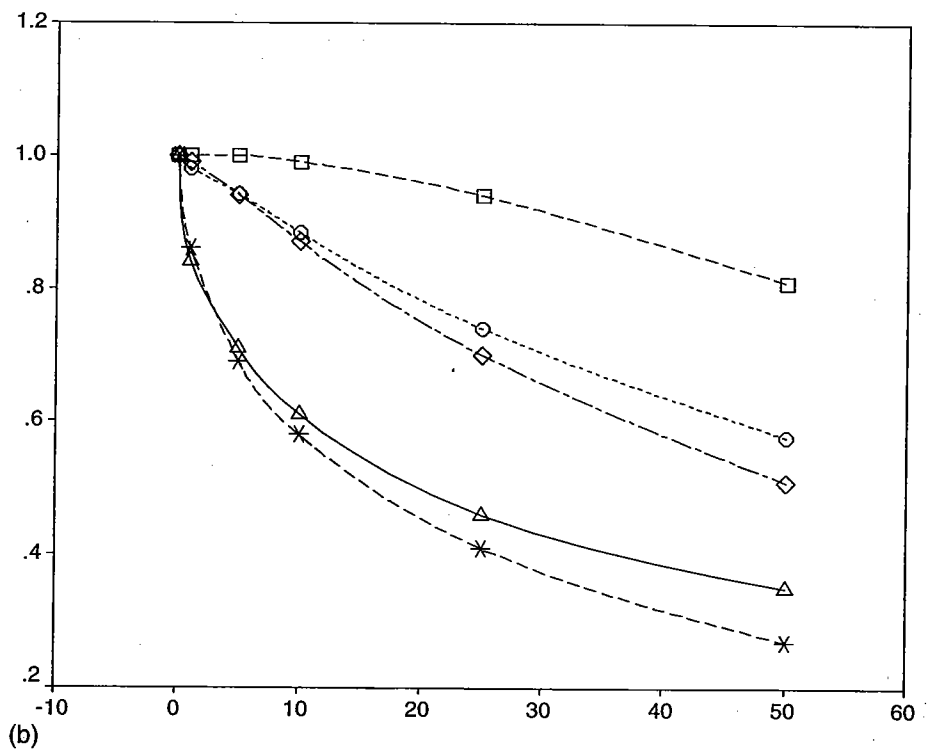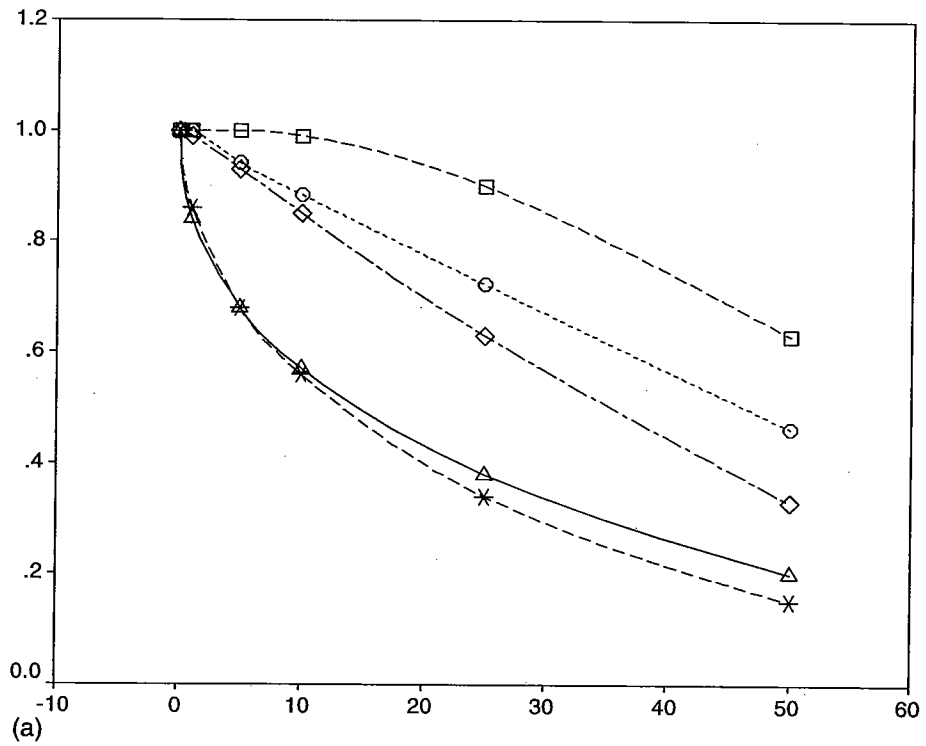| % Error	Degree centrality				Betweenness				Closeness				Eigenvector			
	Node removal	Node addition	Edge removal	Edge addition	Node removal	Node addition	Edge removal	Edge addition	Node removal	Node addition	Edge removal	Edge addition	Node removal	Node addition	Edge removal	Edge addition
Overlap																
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	.84	.85	.82	.82	.84	.84	.81	.81	.84	.85	.82	.82	.84	.85	.82	.81
5	.68	.72	.64	.64	.68	.71	.63	.63	.68	.72	.64	.64	.68	.73	.64	.64
10	.58	.64	.54	.53	.57	.61	.53	.53	.58	.64	.54	.53	.58	.65	.53	.53
25	.39	.51	.37	.37	.38	.46	.36	.36	.39	.51	.37	.37	.38	.54	.37	.37
50	.21	.42	.24	.23	.20	.35	.23	.23	.21	.42	.24	.23	.21	.45	.23	.23
R^2																
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1.00	1.00	.98	.98	1.00	.98	.98	.98	1.00	1.00	.98	.98	1.00	1.00	.98	.98
5	.94	.94	.90	.90	.94	.94	.90	.90	.94	.94	.90	.90	.94	.96	.90	.90
10	.90	.90	.81	.81	.88	.88	.81	.81	.90	.90	.81	.81	.90	.92	.81	.81
25	.74	.81	.59	.59	.72	.74	.58	.58	.74	.81	.59	.59	.74	.83	.59	.59
50	.48	.69	.32	.34	.46	.58	.31	.31	.48	.69	.32	.32	.48	.72	.32	.32

Fig. 2. Scatter plots of average betweenness accuracy as a function of error level for all four kinds of error (limiting to graphs with 100 nodes and 50% density). Each line represents a different accuracy measure. Top line: Top 10%. Second line: R^2. Third line: Top 3. Fourth line: Overlap. Bottom line: Top 1.
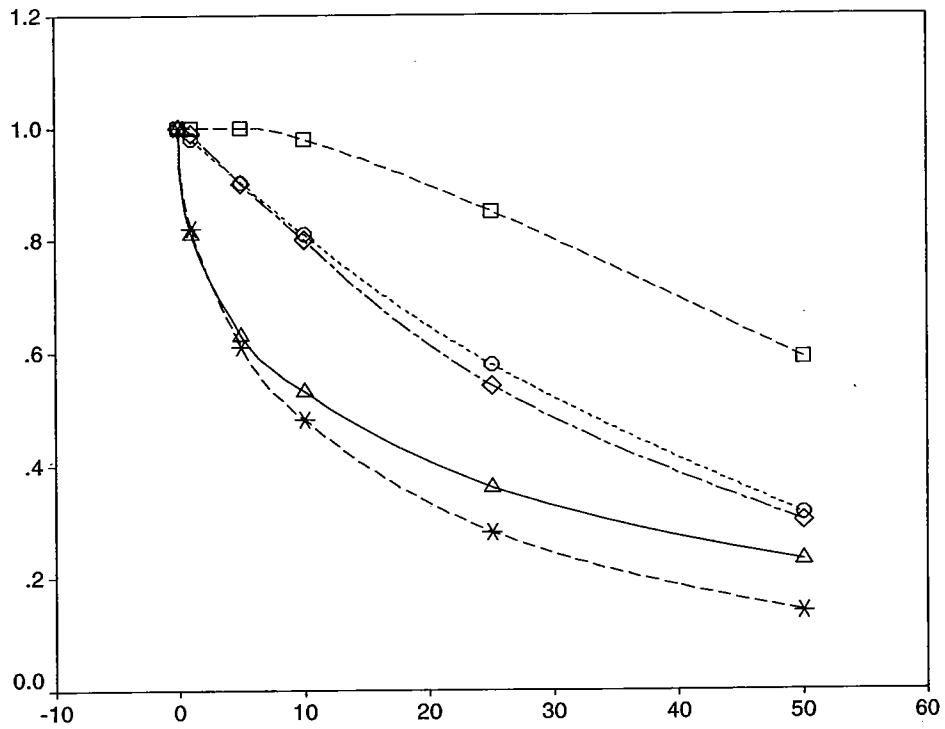
(c)
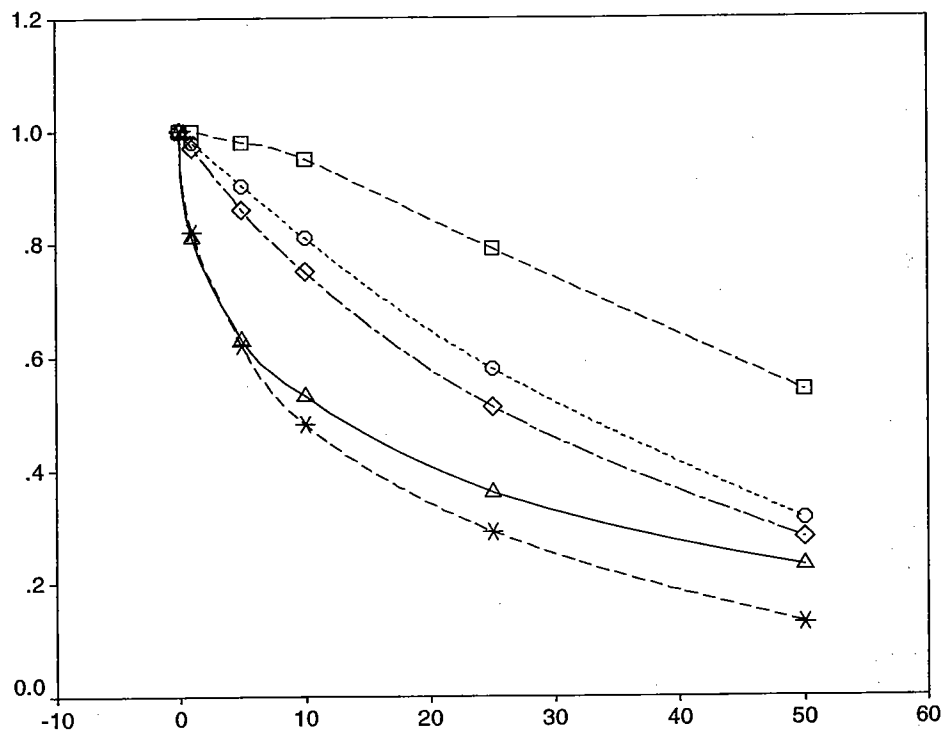
(d)

Fig. 2. (*Continued*).

Table 3

Standardized regression coefficients for density only in regressions of betweenness accuracy measures on density and error level, for graphs of size 100

Error type	Top 1	Top 3	Top 10%	Overlap	R^2
Edge addition	.408	.469	.495	.412	.504
Edge removal	−.267	−.287	−.348	−.269	−.177
Node addition	−.164	−.162	−.219	−.2	−.156
Node removal	.014	.025	.001	.004	.003

Each coefficient is drawn from a separate regression.

192 experimental combinations, and the dependent variables are average accuracy scores. The results indicate that, controlling for error level, density enhances accuracy in the case of edge additions, but reduces accuracy in the case of edge deletion. This moderated effect of density on accuracy is probably due to the number of dyads available to be changed. In the case of edge removal, greater density means more ties available to be removed, making greater changes in the network as a whole and therefore reducing accuracy. In the case of edge addition, greater density means fewer untied dyads available to be tied, and so fewer changes to the network as a whole. Thus, the underlying variable is the absolute amount of change to the network—the Hamming distance between the true and observed networks. A similar effect may occur with node addition. When we insert nodes, we add ties from the new nodes to the old ones and to the other new ones. The number of ties added is a non-linear function of the density of the original network. As a result, denser networks result in disproportionately more ties being added which increases the Hamming distance between true and observed networks.

Similar regressions involving network size (see Table 4) show that network size is only weakly related to accuracy, especially in the case of the r^2 measure, where the effect was essentially zero. For the Top 1 and Top 3 measures, the effect of size is negative. This occurs for strictly structural reasons: as networks get larger, the chances that a particular node occupies the same status slot (such as Top 1 or Top 3) in a perturbed version of the network gets progressively smaller—i.e., the ratio of possible matches to possible non-matches decreases with increasing size. For the Top 10% measure, the effect of size is positive. This is because the Top 10% criterion (the proportion of times that the most central node in the true network is found in the top 10% of the observed network), effectively makes the "strike zone" for correct matches proportional to the network size. As a result, size is positively related to accuracy measured in this way.

Table 4

Standardized regression coefficients for network size in regressions of betweenness accuracy on size, density and error level

Error type	Top 1	Top 3	Top 10	Overlap	R^2
Edge addition	−.166	−.188	.122	−.120	−.067
Edge removal	−.261	−.290	.157	−.220	−.087
Node addition	−.295	−.276	.182	−.253	−.078
Node removal	−.224	−.265	.221	−.194	−.151

Each coefficient is drawn from a separate regression.

4. Summary and conclusion

We have systematically explored the robustness of centrality measures in the face of varying amounts and types of measurement error. A number of conclusions can be drawn. Perhaps the most important is that accuracy not only declines with increasing error, but does so predictably and monotonically. The implication of this finding is that, in principle at least, if one knows the rate and type of error in the data collection process, one can establish error bounds on the metrics constructed from the observed data.

Another key finding was that the four centrality measures considered in this paper are surprisingly similar with respect to pattern and level of robustness. In terms of pattern of robustness – the response curves across experimental conditions – the measures are essentially identical. In terms of overall level of robustness, the measures are extremely similar, with betweenness consistently a hair lower than the other three across all measures of accuracy. This goes against the intuitive notion that global measures of centrality like betweenness would be more sensitive to changes in the network than local measures like degree, since they are potentially affected by any change in the network no matter how distant. It may be that the networks in our study – all of which were random graphs – contained enough redundancy at the path level to counteract the sensitivity of the global measures. If each pair of nodes were connected by multiple geodesic paths, the loss of a few nodes or lines would have much less effect on the global measures.

Another surprising result was that the different types of error had relatively similar effects on centrality robustness. We might have expected a priori that node perturbations would create greater problems for measuring centrality than edge perturbations, because the loss of a node necessarily entails the loss of edges as well. But the results show that type of error makes relatively little difference, and what difference there is runs in the opposite direction: edge perturbation makes more difference. We believe that this result may be limited to random graphs.

Density of the true network tends to reduce accuracy for all kinds of error except edge addition, where it increases accuracy. The contingent effect of density on accuracy is due to the number of dyads available to be changed. In the case of edge removal, greater density means more ties available to be removed, making greater changes in the network as a whole. In the case of edge addition, greater density means fewer untied dyads available to be tied, and so fewer changes to the network as a whole. Thus, the underlying variable is the absolute amount of change to the network—the Hamming distance between the true and observed networks.

One question that we have not specifically addressed is the practical bottom-line—is it reasonable to compute centrality indices when we know that the data contain errors? Based on our results, the answer would seem to be 'yes', as the measures of centrality tested were quite robust under small amounts of error (such as 10% and under). Of course, whether the levels of accuracy are sufficient for any given purpose is difficult to assess since it depends on external factors such as the consequences of error. For example, the results suggest that if our data collection method misses 5% of ties, then the correlation between true and observed centrality will be in the .90s. By social science research standards, the observed score is clearly a superb proxy for the true score. On the other hand, for some applications even this level accuracy may not be sufficient. At that error rate, the probability of correctly

identifying the most central node is around 90%, and the expected overlap in the top 10% is just 67%. One can imagine a situation, perhaps in managing an epidemic, where failing to quarantine or immunize the most central actors could have huge costs in human lives (Borgatti, 2003). In such a case, these results might suggest the need for more accurate methods of data collection.

A crucial limitation of this study is that we have studied only random error on random networks. This is appropriate as a first step in understanding how measurement error affects the calculation of network indices, but it should be clear the results could be quite different for practical settings in which (a) the data collection methodology makes systematic errors (such as more readily losing nodes with low degree), and (b) the networks themselves are not randomly constructed (as we expect for most human networks).

Acknowledgements

References

Bonacich, P., 1972. Factoring and weighing approaches to clique identification. Journal of Mathematical Sociology 2, 113–120.

Bonacich, P., 1987. Power and centrality: a family of measures. American Journal of Sociology 92, 1170–1182.

Borgatti, S., 2003. The key player problem. In: Breiger, R., Carley, K., Pattison, P. (Eds.), Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers. Committee on Human Factors, National Research Council, pp. 241–252.

Brass, D.J., Burkhardt, M.E., 1993. Centrality and power in organizations. In: Nohria, N., Eccles, R. (Eds.), Networks and Organizations: Theory and Practice. Harvard Business School Press, Boston, MA, pp. 191–215.

Carley, K.M., 2003. Dynamic network analysis. In: Breiger, R., Carley, K., Pattison, P. (Eds.), Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers. Comittee on Human Factors, National Research Council, pp. 133–145.

Carley, K.M., Lee, J., Krackhardt, D., 2001. Destabilizing networks. Connections 24 (3), 31–34.

Costenbader, E., Valente, T.W., 2003. The stability of centrality measures when networks are sampled. Social Networks 25, 283–307.

Erdos, P., Renyi, A., 1959. On random graphs. I, Publ. Math. Debrecen 6, 290–291.

Everett, M.G., Borgatti, S.P., 2005. Ego-Network Betweenness. Social Networks. 27 (1), 31–38.

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

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

Marsden, P.V., 1993. The reliability of network density and composition measures. Social Networks 15 (4), 399–423.

Moreno, J.L., 1934. Who Shall Survive?. Nervous and Mental Disease Publishing Company, Washington, DC.

Podolny, Joel M., 1993. A status-based model of market competition. American Journal of Sociology 98 (4), 829–872.

Roethlisberger, F., Dickson, W., 1939. Management and the Worker. Cambridge University Press, Cambridge.

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge.