Scientific Research Publishing

Confidence Intervals for Assessing Sizes of Social Network Centralities

Dawn Iacobucci[1]*, Rebecca McBride[2], Deidre L. Popovich[3], Maria Rouziou[1]

[1]Vanderbilt University, Nashville, TN, USA
[2]Finance and Administration Development, Deloitte, New York, NY, USA
[3]Texas Tech University, Lubbock, TX, USA
Email: *dawn.iacobucci@owen.vanderbilt.edu, becanya@hotmail.com, deidre.popovich@ttu.edu,
maria.rouziou@owen.vanderbilt.edu

Abstract

This research uses random networks as benchmarks for inferential tests of network structures. Specifically, we develop formulas for expected values and confidence intervals for four frequently employed social network centrality indices. The first study begins with analyses of stylized networks, which are then perturbed with increasing levels of random noise. When the indices achieve their values for fully random networks, the indices reveal systematic relationships that generalize across network forms. The second study then delves into the relationships between numbers of actors in a network and the density of a network for each of the centrality indices. In doing so, expected values are easily calculated, which in turn enable chi-square tests of network structure. Furthermore, confidence intervals are developed to facilitate a network analyst's understanding as to which patterns in the data are merely random, versus which are structurally significantly distinct.

Keywords

Centrality, Degree, Closeness, Betweenness, Eigenvector Centrality, Social Networks

1. Introduction

Many social network analyses begin with identifying actors with high centrality. Depending on the nature of the actors and the ties that link them, highly central actors may be actors with power or prestige, expected to be influential in the network ([1] Bonacich, 1987; [2] Mizruchi and Potts, 1998). Numerous centrality indices are available to characterize actors' positions amidst the structural ties in social networks, including degree, closeness, betweenness, and eigenvector

centralities, among others, each with a distinct conceptualization ([3] Freeman, 1978; [4] Bonacich, 1972).

It is not unusual for social network scholars to obtain centrality scores to understand a network structure in a descriptive sense, and perhaps to use the centrality indices as input into subsequent analyses to study associations with actor attributes or behaviors of the actors or network. Descriptive statistics on networks are certainly very informative, yet even more information may be obtained via inferential methods with statistical methods that allow testing hypotheses about network structure. For example, social network analysts might pose questions such as whether a set of actors are significantly different from others, to help formulate conclusions about patterns that appear different, making those assessments with greater confidence. This study is intended to contribute to the literature on inferential analyses of social network data. We take a relatively novel and straightforward approach to obtain the statistics, and we believe it is at least as important that this approach should be easily implemented by users for their own social networks.

In building toward these statistics, we begin in Study 1 by examining the descriptive statistics for four popular centrality indices across several stylized networks. (These particular centrality indices are the four that are most frequently implemented in network texts and software). The network prototypes are drawn from the social networks literature which has frequently used clean structures to illustrate various characteristics of networks, from centrality features to cliques and equivalence patterns. We then add error to the networks so the structures of ties might resemble real data more closely than their stylized forms, and we compare the descriptive statistics as more noise is introduced into the network. The results show that as the networks become more errorful, their statistical profiles become similar, regardless of the initial structure of the network.

Given the convergence in similar centrality indices, we conduct a complementary investigation in Study 2, in which we focus on fully random networks. Networks of random ties can serve as a benchmark against which real network data may be compared to assess the extent to which the observed network structures are *real* patterns and phenomena versus merely those that would be observed in networks of random collections of ties. We use this information to develop formula to test hypotheses about centrality scores. Specifically, we derive *expected values* for centrality indices and compare those against observed network centralities in a *chi-square* test, with follow-up *z*-tests. Additionally, we create *confidence intervals* to form bands within which observed network ties may be said to be expected at the rate of a random network, and beyond which, in either direction, higher or lower real centralities may be said to reflect truer, non-random network structure in the social network data.

It is our hope that this research and these tools might be useful in assisting network researchers in determining whether their observed sets of network patterns are reliable and statistically significant network structures, above and beyond

random collections of ties. We derive these expected values, the statistical tests, and the confidence intervals for each of four popular centrality indices: degree, closeness, betweenness, and eigenvector centralities.

The paper is organized as follows. We first briefly review the centrality indices that comprise the focus of this research. In Study 1, we examine the centrality indices on three stylized networks, and observe the centrality profiles as random errors are introduced into the networks. In Study 2, we consider results for wholly random networks with varying numbers of actors and network densities. We use the results to derive more general formula, for any number of actors and density, to obtain expected values and confidence intervals for each of the four centralities.

2. Four Focal Centrality Indices

This research focuses on four centrality indices: degree, closeness, betweenness, and eigenvector centrality. Degree reflects overall volumes of ties, closeness captures the extent to which the relational ties travel via few "degrees of separation", and betweenness highlights those actors through whom much of the rest of the network is interconnected. Eigenvector centrality is a weighted function that incorporates information about an actor's connections with other actors who may themselves be highly central. Scholars have proffered additional centrality measures, but these four are prevalent across texts on social network analysis (e.g., [5] Knoke and Yang, 2007; [6] Wasserman and Faust, 1994) and across social network analysis packages (e.g., UCINet, NetMiner, LibSNA, NodeXL, even Mathematica; each network package may provide additional indices, but such indices do not appear consistently across network packages, unlike the four core centralities in our studies). Next, we briefly review each index.

Degree. The notion of an actor degree is intuitively understood as capturing the volume of interconnections ([3] Freeman, 1978; [7] Bolland, 1988; [5] Knoke and Yang, 2007; [8] Lu, Chen, Ren, Zhang, Zhang, and Zhou, 2016; [9] Opsahl, Agneessens, and Skvoretz, 2010; [10] Rothenberg *et al.*, 1995; [6] Wasserman and Faust, 1994). For a $g \times g$ sociomatrix on g actors, $X = \{x_{ij}\}$, with actors in rows $i = 1, 2, \cdots, g$ sending ties to the actors in columns $j = 1, 2, \cdots, g$, the in-degree is defined as the column sum and the out-degree as the row sum. For binary ties, the in-degree and out-degrees for actor or node "*i*" are:

$$C_{D-in}(i) = \sum_{j=1, i \neq j}^{g} x_{ji} \quad \text{and} \quad C_{D-out}(i) = \sum_{j=1, i \neq j}^{g} x_{ij} .$$

Actor degrees are normed as

$$C_D'(i) = \frac{C_D(i)}{g-1} ,$$

for $(g-1)$ being the maximum number of links for an actor to the others in the network.

Closeness. An actor's closeness is "based upon the degree to which [an actor] is close to all other [actors] in the [network]" ([3] Freeman, 1978, p. 224).

Closeness uses the geodesics, or shortest paths between actors ([11] Brandes, Borgatti, and Freeman, 2016; [3] Freeman, 1978; [12] Kim, Hyun, and Kim, 2015; [13] Smith and Moody, 2013; [14] Stephenson and Zelen, 1989). Given d_{ij} as the distance or number of edges in the geodesic linking actors i and j, or the length of the shortest tie that connects them, closeness is the inverse of the sum of the geodesic distances, defined for binary and mutual or symmetric ties (*i.e.*, $X = X'$) as follows:

$$C_C(i) = \frac{1}{\sum_{j=1, i \neq j}^{g} d_{ij}}.$$

At most, one actor may be as far as $(g-1)$ steps from another, so closeness indices are normed: $C_C'(i) = (g-1)\left[C_C(i)\right]$.

Betweenness. Betweenness is "based upon the frequency with which [an actor] falls between pairs of other [actors] on the shortest or geodesic paths connecting them" ([3] Freeman, 1978, p. 221; [15] Costenbader and Valente, 2003; [2] Mizruchi and Potts, 1998; [16] Riondato and Kornaropoulos, 2016; [17] Venkataramani, Zhou, Wang, Liao, and Shi, 2016; [18] Xiao, Wu, and Huang, 2014; [19] Zemljič and Hlebec, 2005). With g_{ijk} representing the number of geodesics linking j and k that contain actor i, the betweenness indices on binary and symmetric ties are defined:

$$C_B(i) = \sum_{j<k}^{g} \frac{g_{ijk}}{g_{jk}}.$$

The maximum value is $\left[(g-1)(g-2)\right]/2$, so this centrality is normed:

$$C_B'(i) = \frac{2C_B(i)}{(g-1)(g-2)} = \frac{2C_B(i)}{g^2 - 3g + 2}.$$

Eigenvector. [4] Bonacich (1972), building on [20] Katz (1953), proposed that the first eigenvector (corresponding to the largest eigenvalue) of a sociomatrix could serve as a centrality measure that would capture patterns of direct and indirect connections ([21] Bonacich, 2007; [22] Bonacich and Lloyd, 2015; [23] Borgatti, Carley, and Krackhardt, 2006; [24] Li, Liu, Jiang, and Liu, 2016). The index captures a weighting of ties such that actors would have a larger eigenvector index to the extent that they were connected to other actors who themselves are highly inter-connected. For the sociomatrix X, the eigenvector, v, is derived from the familiar equation: $Xv = \lambda v$. The eigenvector score for actor i is $C_{EV}(i)$, a weighted function of the statuses of the other actors to whom actor i is connected: $C_{EV}(i) = x_{1i}v_1 + x_{2i}v_2 + \cdots + x_{gi}v_g$.

Various research articles have certainly considered alternative centrality indices, such as information centrality ([10] Rothenberg *et al.*, 1995; [14] Stephenson and Zelen, 1989), or indices for directed graphs ([25] Freeman, Borgatti, and White, 1991; [26] White and Borgatti 1994). However, the four indices we have selected—degree, closeness, betweenness, and eigenvectors—seem most transferable in their general use across network textbooks, research articles, and available software packages.

Next, in Study 1, we examine the empirical performance of these four centrality indices on small social networks that have exemplar, stylized network structures drawn from the literature. We will then perturb the clean, prototype structures by adding random error to the network ties and observe the effects on the four centrality indices.

3. Study 1: Performance of the Centralities on Stylized Networks

To understand the nature of the differences among the centrality indices, it should be useful to begin with clean networks, simple and classic in appearance. Figure 1 depicts social networks with prototypical structures that have been used to inform the conceptual development of many social networks analytics, from centrality indices to definitions of cliques and stochastic equivalence. First is the "hierarchy" (or "fork" or "tree" per [3] Freeman, 1978, p. 233, [6] Wasserman and Faust, 1994, p. 468), a structure frequently used to depict structures of organizations, for example. In terms of network centralities, the top player may have the largest degree, and the middle layer may have the largest betweenness indices. Next is the "star" or "wheel" structure with one highly central actor whose ties emanate out to other actors, who are not themselves connected (cf., [3] Freeman, 1978, pp. 219, 233; [6] Wasserman and Faust, 1994, p. 171). The star network has regular or uniform degrees with the exception of the central-most actor, who is both closest to all the others, and through whom the others must traverse to reach other actors in the network. Thus, for example, closeness indices should be reasonably high for these actors. The final figure depicts a "core and periphery" structure in which a subset of actors within the network are highly interconnected (at the extreme forming a clique), and in which a second set of actors is connected to the first, but not as completely linked to those in the first set nor to each other (cf., [27] Borgatti and Everett, 1999; [28] Mizruchi and Potts, 1998, p. 357).

The intention in selecting these particular stylized networks was to represent some variability across structural properties that might be reflected better by one or more of the centrality indices. Indeed given these structures, it would not seem unreasonable to anticipate that some centrality indices may be more sensitive to certain elements of different network structure. For example, one might expect the core-periphery network to have high degrees, and high closeness,
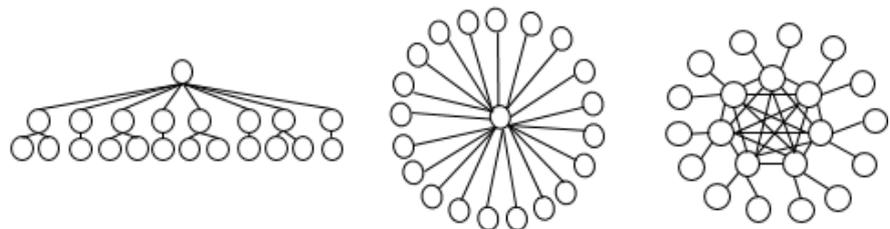
Figure 1. Three stylized networks: Hierarchy, star, core-periphery.

whereas the hierarchy might yield greater betweenness indices. Nevertheless, before proceeding, it is important to note that even if one might have preferred that a different selection of network structures be included for examination, the particular structures will not matter soon, as we add random noise, which will become clear shortly.

3.1. Means of the Four Centralities, for Each Network

To examine whether these relationships hold, we analyzed each network to obtain all four sets of centralities. (For simplicity, we constructed the adjacency matrices to be binary and symmetric). The means of each centrality computed across the actors are presented in Table 1, for each network. The pattern of results are essentially as anticipated, e.g., betweenness scores being higher for the hierarchy than for the star or core-periphery, both of which have more direct ties. Similarly, the core-periphery had larger degrees, and so forth. These results are not critically of interest except that they serve as a foundation to compare results when we add noise. We do so next.

These stylized networks are exemplar structures that should epitomize network patterns for which one type of centrality index would be optimal to use. Yet the networks are so simple and clean that they do not seem particularly representative of real network data. Thus, we build further, adding noise to these stylized networks.

Noisy or errorful networks have been used to study numerous network phenomena ([7] Bolland 1988; [15] Costenbader and Valente 2003). For example, [28] Watts and Strogatz (1998, p. 441) used structures of random networks as a benchmark against which to compare an observed network to verify the existence of a network clustering characteristic beyond chance. [15] Costenbader and Valente (2003) sampled from eight real data sets, calculated 11 centrality indices, and tested recovery to the centralities of the full, intact networks, showing that in-degrees and eigenvector scores were relatively stable until sample size proportions dropped below 40% - 50%, whereas closeness and betweenness indices were diminished even when larger portions of the data were retained. [19] Zemljič and Hlebec (2005) obtained real network data with different levels of measurement (e.g., binary, categorical, ratio from continuous line generation, etc.) and found greater test-retest reliability for degree and closeness centralities than for betweenness indices, particularly for more dense networks. [23] Borgatti,

Table 1. Mean centrality indices per stylized network.

Centrality Index	Network Structure		
	Hierarchy	Star	Core-Periphery
Degree	0.095	0.095	0.167
Closeness	0.360	0.536	0.470
Betweenness	0.100	0.048	0.065
Eigenvector	0.173	0.184	0.162

Carley, and Krackhardt (2006) studied networks of varying size (g = 10 to 100) and densities (1% to 90%) to understand the stability of network indices as network data are sampled, simulated by the addition or deletion of nodes or edges. The perturbations were more disruptive to network recovery than sampling variations on the nodes. [13] Smith and Moody (2013) sampled nodes from 12 real data sets to study missing data, and found greater robustness for in-degrees and closeness than betweenness centralities, with large and centralized networks displaying the least bias in the estimates of the remaining network data.

Real datasets can serve as an acceptable truism, against which the effects of sampling and adding noise may be compared. However, in real data, the extent to which betweenness or closeness, say, should reflect elements of the true, underlying network structure is unknown. Hence, we will add noise to our previously analyzed stylized networks to gauge the sensitivities of the centrality indices, having begun from bases with pure, known network patterns.

3.2. Adding Noise to the Stylized Networks

We continue this investigation by perturbing the network structures by adding random error from a uniform distribution. Specifically, for each network structure, and each cell in the sociomatrix, a 0 (1) was changed to a 1 (0) with probabilities that ranged from 0.0 to 0.5. For example, for a probability of 0.2, on average, 80% of the ties in the network remained the same, with 20% reversals. Changes were made to the upper triangle of the matrix, and then copied to the lower triangle so as to maintain symmetry. Once the sociomatrix was revised, the centrality indices were calculated. This process was repeated 100 times for each combination of network structure and level of error.

The results are presented in Figure 2. The results for zero error—the left-most points in each plot—represent the perfectly stylized networks of Figure 1, with means that had been conveyed in Table 1. What is new in Figure 2 are the means for 0.1 through 0.5 error levels, which allow us to observe the mean centralities migrating from idealized network forms to random collections of ties lacking systematic forms.

In the results, note that the mean centrality indices for each network begin at slightly different values (per Table 1), where each network is in its purest form without error. From there, as error is added and the network structures perturbed, the inherent structures of the networks (e.g., hierarchy, star, etc.) have less of an impact in determining the centrality indices because more random ties are contributing as well.

As error is introduced, the average degree centrality increases for each network, as does the average closeness centrality. The average betweenness centrality decreases slightly (not having far to drop from low initial values). The eigenvector indices remain stable, increasing only a modest amount (e.g., from 0.162 to 0.214 for the core-periphery network).

As the level of noise added approaches the level of 50%, the centrality indices
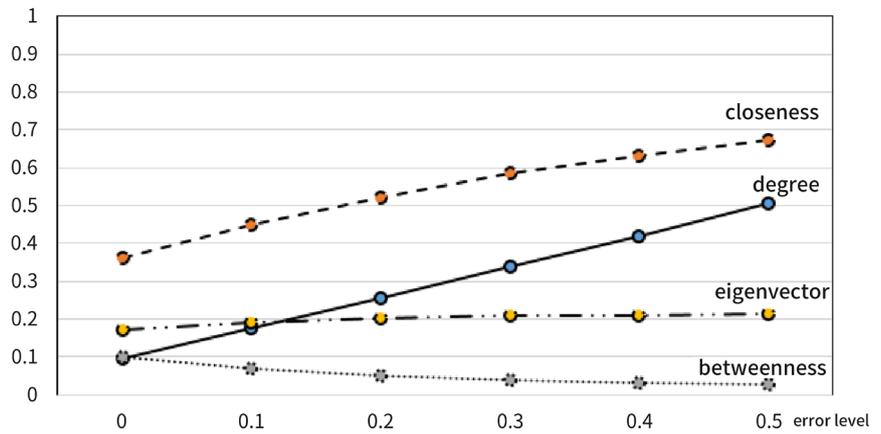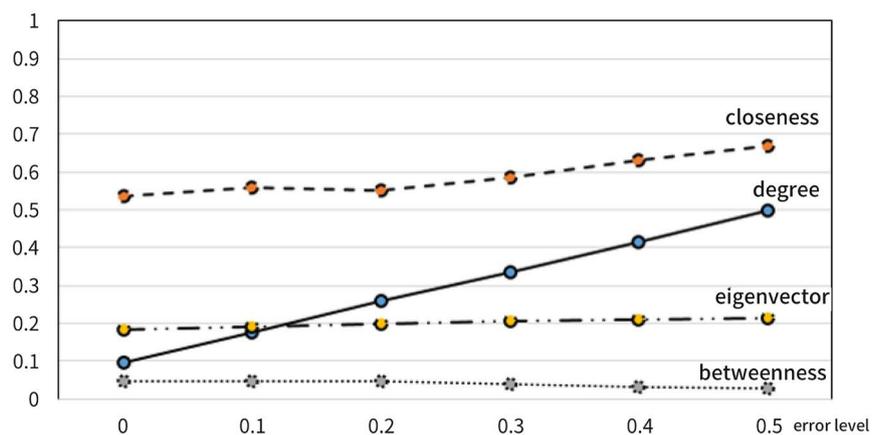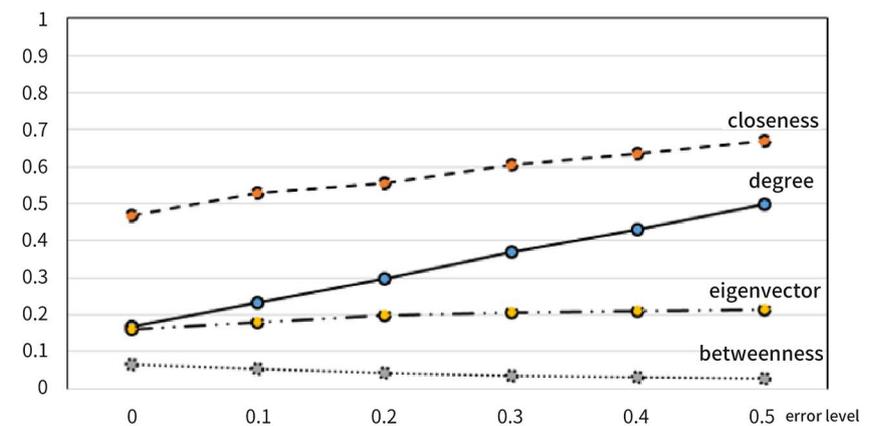
Figure 2. Comparative descriptives on "Stylized + Noise" centrality indices. (a) Mean centralities with varying errors—hierarchy network; (b) Means of centralities with varying errors—star network; (c) Means of centralities with varying errors—core-periphery network.

converge. This result may be anticipated by some, or may certainly be easy to understand in hindsight. Specifically, the indices should converge regardless of

which of the three starting networks on which it is being calculated, because as error is added to the 50% level, the network is at its noisiest, essentially a random network, with none of the original inherent structure remaining. The hierarchy network, for example, began in its clean form depicted in Figure 1. With the noise level at 0.5, each tie, present (absent), had a 50:50 chance of remaining present (absent) or being modified to a state of being absent (present). The values to which all networks converge are: 0.50 for degree, 0.67 for closeness, 0.03 for betweenness, and 0.21 for eigenvectors. That is, the right-most values in each panel, regardless of the initial network structure, all share the values of 0.50, 0.67, 0.03, and 0.21. These values are therefore "expected values" for networks comprised of random collections of ties. For real network data, one would hope that the structures would yield centrality indices (significantly) different from (higher or lower) these base expected values.

This finding makes sense because the networks at the right in each panel of Figure 2 are all random networks. The networks at the left had certain structures (hierarchy, star, core-periphery) but those structures have been replaced at the right in the plots with networks comprised of random ties, heads or tails as to whether a tie connected any given dyad of actors. Thus the random networks to the right in the first panel (that had begun as hierarchy networks) are simply a random sample from a population network without structure, that is, a random network. The random networks to the right in the second and third panels similarly represent random samples from a population network with no inherent structure. Given that all three panels show random networks at the right (where the error levels are 0.5), each of the four centrality indices converge (across panels) such that closeness, for example, on the random networks in the first panel are, within sampling, the same values as the closeness indices for the random networks in the second and third panels. In the study that follows, we pursue this observation about random networks in greater detail, to create a baseline for the centralities indices. Once we have a baseline of what value a centrality index may be expected, we can compare observed centralities that have been calculated from networks that have some non-random structure. The comparison will allow summary statements as to the extent to which a network varies from randomness.

Therefore, next we shall show that in their random states, the four centralities may be derived as a function of the size of the network or the number of actors, *g*, and the network's density, the proportion of extant ties. In the section that follows, we show this relationship to be precisely true for degree centrality, and approximately true for closeness, betweenness, and eigenvector centralities.

4. Study 2: Fully Random Networks

In this next investigation, we consider the purely random network form to understand what the centrality indices may be measuring in such contexts. We will derive the expected values for degree centralities analytically, and empirically

calculate the expected values for closeness, betweenness, and eigenvector centralities. In addition, we will provide the formulas for confidence intervals for each of the centrality indices.

Figure 3 shows the progression of covering a continuum from stylized to random networks. Specifically, whereas Study 1 considered clean, stylized networks, as well as results examining the effect of added noise, we turn now to the scenario of fully random networks; *i.e.*, networks that have been generated as collections of random ties.

1) At the left are the stylized networks presented in Figure 1 and analyzed in Table 1 and the left-most points of Figure 2.

2) The scenario of "Stylized + Noise" were networks that were analyzed and mean centralities presented in the panels of Figure 2 for the 0.1 to 0.5 random error levels.

3) The fully random networks will be analyzed in Figure 4, to be presented shortly.

To proceed in the creation of the random networks, we varied networks in size, g = 50, 100, 150 actors, and densities = 0.1, 0.3, 0.5, 0.7. For each combination of parameters, e.g., g = 50 and density = 0.7, we created a network with those specifications and proceeded to calculate the four centrality indices, noting their descriptive statistics and correlations. This process was repeated 100 times for each combination.

Let us consider first the mean centralities, presented in Figure 4 for each of g = 50, 100, 150 actors. For degree centralities, regardless of the number of actors in the network, g, the values start at 0.10 (for 10% density), and achieve mean values of 0.7 (for 70% density). We derive the relationship between the expected degree value and density analytically shortly, and like many things, the relational form will appear obvious in hindsight.

For the closeness centrality indices in a 10% density network, the values are 0.41, 0.45, and 0.48 for g = 50, 100, 150 respectively, whereas in a 70% density network, the values are 0.77, for all g. We will formulate the expected values for closeness shortly.

The results are different for the betweenness and eigenvector centralities. Both remain relatively constant regardless of the densities of the network, or network size, g. The average betweenness scores range from 0.03 (for 10% density) to 0.01 (for 70% density) for g = 50, 0.01 to 0.003 for g = 100, and 0.01 to 0.002 for g =
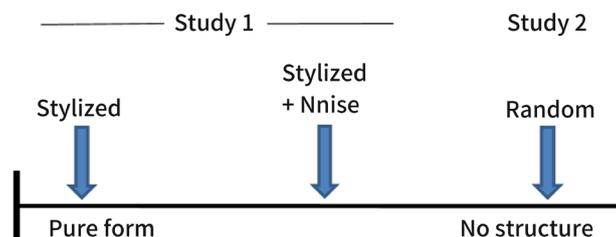
Figure 3. Classes of networks.

(a)

(b)

(c)
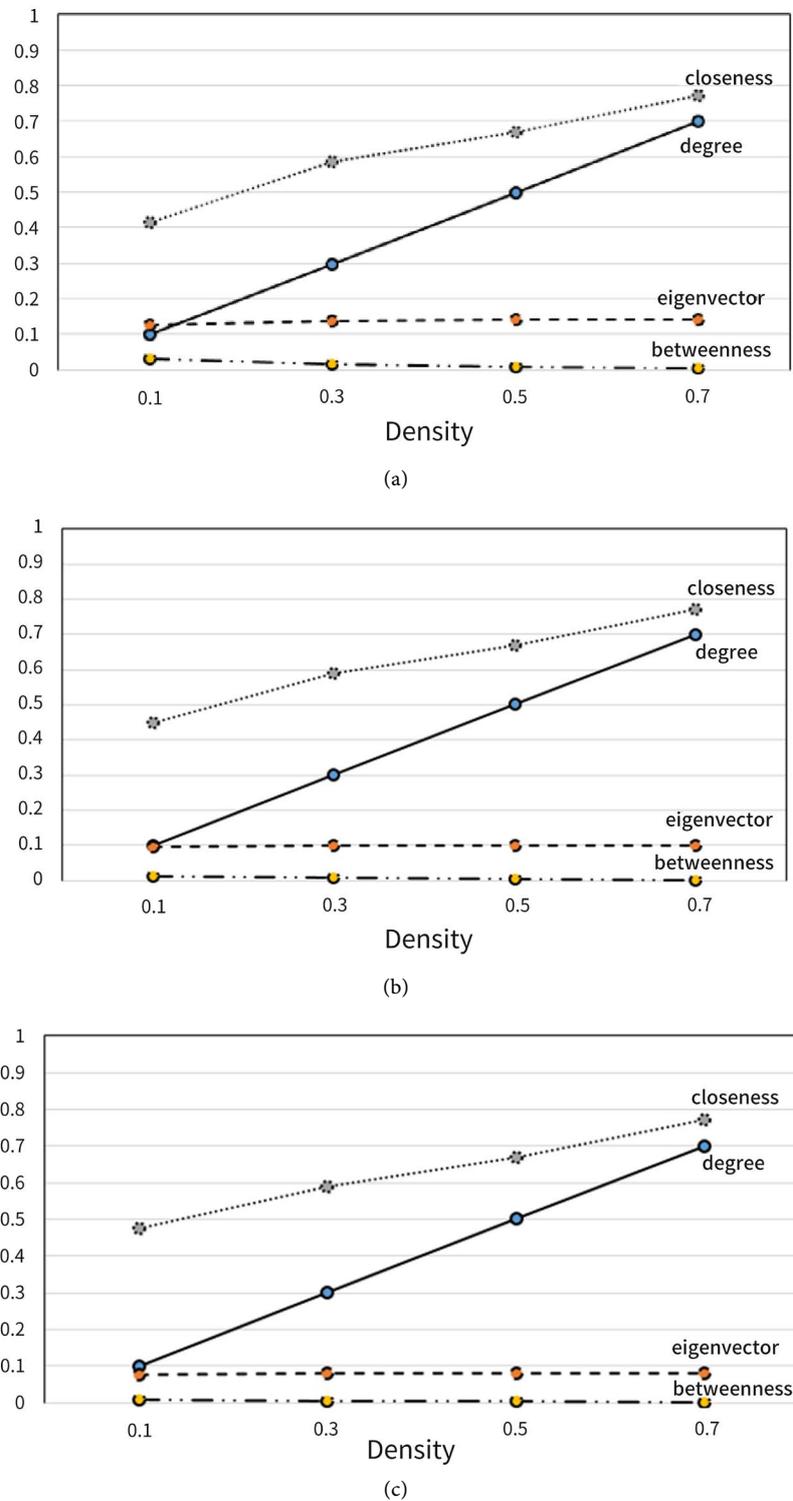
Figure 4. Mean centralities on random networks. (a) 50 actors; (b) 100 actors; (c) 150 actors.

150. The changes in the average eigenvector scores are similarly modest, increasing from 0.13 for 10% density to 0.14 for *g* = 50, and 0.095 to 0.099 for *g* = 100 actors, and 0.079 to 0.082 for *g* = 150 actors. It is perhaps sensible that the

betweenness and eigenvector centralities are rather insensitive to differences in densities, even if neither is explicitly normed to adjust for the prevalence of ties. Betweenness values reflect the inter-connectivity, and eigenvectors the direct and indirect ties, both doing so regardless of the overall volume of ties, indeed both are presumably stabilizing with increased density as, by definition, the direct connections increase, therefore leaving fewer indirect paths remaining, which affect both betweenness and eigenvector scores. Note that by comparison, these results almost suggests that social network analysts should introduce a normative adjustment for degree and closeness centralities to account for density (not just network size) so as to tease out that confound from indices intended to reflect actors' patterns of connections.

Across the panels in Figure 4, it might seem odd at first that the means for the betweenness and eigenvector centralities are flat, whereas degree and closeness climb. Yet upon further consideration, the result makes sense. Let's take degree and closeness first. For degree centralities, as density increases, the number of ties has increased for a fixed number of actors, hence the additional ties become distributed in some fashion across the actors, thereby boosting the degrees. Similarly, closeness increases because with the additional ties, it becomes ever more likely that any dyad of actors becomes directly connected, thereby reducing the number of links of separation, thereby strengthening closeness. In contrast, betweenness centralities stay relatively constant, because these centralities reflect a particular structural formation, not simply the presence of a greater (or lesser) number of ties. If an actor lies between two groups, the actor is still the bridge through which the other connections must pass. Certainly as ties are added (*i.e.*, density increases), one would expect there to be more direct connections between the two groups that had previously only the one actor serving as an intermediary conduit, so that is presumably a matter of proportions (*i.e.*, size of networks, numbers of highly between actors, etc.). (It should be mentioned again that the betweenness centralities did indeed decline, albeit very modestly, from 0.03 to 0.006 for $g = 50$ actors, 0.01 to 0.003 for $g = 100$ actors, and 0.007 to 0.002 for $g = 150$ actors. That is, the index behaved in the direction expected, it simply had not far to fall). Lastly, the eigenvector centralities appear flat, but this too is sensible in that the eigenvectors capture more explicitly than the other three centralities both direct and indirect ties, though here too, while the increases are so slight that they are difficult to discern in the graphs, the eigenvector centrality values increased modestly presumably reflecting the influx of some additional direct ties (*i.e.*, from 0.13 to 0.14 for $g = 50$ actors, 0.09 to 0.10 for $g = 100$ actors, and 0.079 to 0.081 for $g = 150$ actors).

4.1. Benchmarks: Expected Values

These various results on fully random networks can be used to derive baselines for the purposes of comparing real network structures and determining the extent of the validity of the inherent patterns in the network ties. For example, for

the mean centralities depicted in Figure 4, we note that the points and seeming constancies across the panels (depicting varying network size g's) are meaningful. For symmetric, binary sociomatrices, we can easily derive an expected value for degree centralities for a network, in the mathematical statistical sense, as a function of its size (g) and density. Consider the formula for density, defined as the proportion of existing ties relative to all that might exist:

$$\text{Density} = \frac{\#\text{ties}}{\#\text{possible}} = \frac{\sum_{i=1}^{g}\left(\sum_{j=1,\neq i}^{g} x_{ij}\right)}{g(g-1)}$$

Given the formula for a degree centrality is:

$$C_D(i) = \sum_{j=1,\neq i}^{g} x_{ij}$$

Then density may also be written as:

$$\text{Density} = \frac{\sum_{i=1}^{g} C_D(i)}{g(g-1)}.$$

Furthermore, note the mean degree centrality is:

$$\overline{C_D} = \frac{\sum_{i=1}^{g} C_D(i)}{g}$$

so:

$$\text{Density} = \frac{\overline{C_D}}{g-1}.$$

Thus, if we have the mean centrality, dividing it by $(g-1)$ yields density, or if we have a network's density, we can multiply it by $(g-1)$ to obtain the mean centrality. Note also that given the normed degree centrality is:

$$C'_D(i) = \frac{C_D(i)}{g-1}$$

the mean normed centrality would be:

$$\overline{C'_D} = \frac{\sum_{i=1}^{g} C'_D(i)}{g}$$

So density may also be written as a function of the mean normed centrality. Specifically:

$$\text{Density} = \overline{C'_D}.$$

Thus, if we create a random network with density 0.7, say, then the mean of the normed degree centrality indices will be 0.7 also. In Figure 4, for each $g = 50$, 100, 150, these relationships hold. That is, for random networks with densities of 0.1, 0.3, 0.5, and 0.7, the mean normed degree centralities were also 0.1, 0.3, 0.5, and 0.7, respectively.

As just previewed, now that this relationship has been established, its nature is rather intuitive—if the overall density is 0.7, then on average, one would expect a

degree centrality to be 0.7 if the network was a manifestation of only random sets of ties. If the network were real data depicting real network structures, presumably the set of actor degrees would vary from 0.7, some being lower and some higher as ties cluster around some actors but not others.

4.2. The Chi-Square Test on Degree Centralities

Next we create a chi-square calculation to compare real network data to a random network to highlight the structural elements of the real network that is not shared by a random network (cf., [28] Watts and Strogatz, 1998). If centrality indices had known distributions, and were not, for example, dependent upon structures of non-independent tie configurations, then a network researcher could follow the usual statistical procedure of calculating a mean and standard deviation from the data to test a hypothetical value (or conduct bootstrapping to build an empirical distribution). However, given the lack of independence, the typical procedure is not directly applicable, yet the logic is still useful. Specifically, the random ties have no such structural connections, they are by definition independent, so they may serve as the effective statistical distribution against which real data may be compared to determine whether the real data are different from the point of comparison, that is, structures of random networks.

For example, let us begin with the case for degree centralities. Upon receipt of network data, a network modeler knows immediately the size of the network, g, and can easily obtain the density. From that, as just shown, the expected value of the normed degree centrality would be the density value (*i.e.*, $\overline{C'_D} = \text{Density}$), and the expected value of the raw (non-normed) degree centrality would be: $\overline{C_D} = \text{Density}(g-1)$. That is, if the network showed no particular structure varying across the actors, then each actor would have a degree centrality of approximately $\overline{C_D}$.

These expected values can be used in the familiar chi-square test:

$$X^2_{(g-1)} = \sum_{i=1}^{g} \frac{(o_i - e_i)^2}{e_i}$$

where, $o_i = C_D(i)$, $e_i = \overline{C_D}$, and X^2 would be distributed as (tested for significance against) the χ^2 on $(g-1)$ degrees of freedom. This statistical distribution is applicable given that the ties are binary, distributed Bernoulli individually, summed to binomial and approximated by the normal distribution (the sum of squares being distributed chi-square; [29] Hogg, Tanis, and Zimmerman 2015).

As an example, consider the core-periphery network depicted in **Figure 1**. It has 35 ties (70 1's in its corresponding sociomatrix) and a density of 0.17, so the e_i's would be $0.17(g-1) = 0.17(20) = 3.40$. The o_i's seen in **Figure 1** are (in some order): 14 (1)'s and 7 (8)'s. The

$$X^2 = 14\left[\frac{(1-3.40)^2}{3.40}\right] + 7\left[\frac{(8-3.40)^2}{3.40}\right] = 14(1.694) + 7(6.224) = 67.282$$

compared to χ^2 on (21 - 1) degrees of freedom, yielding $p < 0.0001$, indicating that the actors' degree centralities indeed vary significantly from a uniformly distributed expected degree as if the ties were random.

Given that chi-squares are most applicable to frequencies, the raw, non-normed degree centralities should be used, as in the example just shown. If a network researcher wished to work with the normed degrees, it is easy to show that the X^2 would merely need to be scaled up by multiplying the normed X^2 by $(g-1)$, obviously to cancel the effect of the norming having previously divided by $(g-1)$. That is, $X^2 = (g-1) X^2_{\text{normed}}$.

4.3. Follow-Up Tests to the Chi-Square

Given that the chi-square statistic is comprised of the sum of squared elements each of which is distributed as a z-statistic ([30] Haberman 1973), we may also use the micro, actor-by-actor level of information as well. Specifically, the standardized residuals, $z = \dfrac{o_i - e_i}{\sqrt{e_i}}$ (each piece to be squared in the chi-square formula above) may be compared to a z-distribution, e.g., for 95% confidence level at ±1.96. To continue with the core-periphery example just analyzed, any of the actors with a degree centrality of "1" (namely those in the periphery) would be deemed not significantly different from what a random network would yield, because

$$z = \frac{1 - 3.4}{\sqrt{3.4}} = -1.30 > -1.96,$$

whereas the actors with degrees of "8" (that is, those in the core) have centralities that significantly exceed those which would result from a random network:

$$z = \frac{8 - 3.4}{\sqrt{3.4}} = 2.49,$$

which exceeds 1.96. Obviously real networks will have finer gradations of degree centralities, and each observed degree centrality value can be tested in this manner.

4.4. Expected Values for Closeness, Betweenness, and Eigenvector Centralities

Expected values are more challenging to derive analytically for closeness, betweenness, and eigenvector centralities, however, they are easily obtained empirically, through the generation of numerous random networks, for fixed g's and densities. To do so, we generated 100 random networks each for 15 levels of varying g (10, 20, 30, ..., 150) and 9 levels of density (0.1, 0.2, ..., 0.9). The means for the standardized degree, closeness, betweenness, and eigenvector centralities, as well as their standard deviations, were obtained.

Rather than presenting 15 × 9 tables of reference values, those values were submitted to regressions to replicate the tabled findings and also to allow for es-

timates of expected centralities for g's or densities not tabled, such as $g = 27$, or 134, etc. Thus we used the results as a database in which we regressed the average degree (or closeness, or betweenness, or eigenvector) centrality against the predictor factors of g and density. The resulting equations appear in Table 2. Note that for most predictions, we added another predictor, the natural log of the number of actors, because the regressions without these terms had shown relatively large residuals for extreme values of the predictors. (The number of actors and its log were not significant contributors to the predictions for degree centralities, hence they are not included in the equations).

It is not surprising that the R^2 for degree centrality is nearly perfect, given the analytical solution shown previously (indeed this empirical approach was not necessary for degree centralities, given the explicit analytical solution just presented). The R^2 for eigenvector and betweenness centralities are high enough to suggest that the equations can be useful.

The R^2 for the closeness centralities is very weak, so those forecasts should be considered very approximate. We sought better predictive models for closeness, and obtained increases in R^2 to levels of 0.3 and 0.4, but the models seemed convoluted, e.g., adding predictive terms such as g^2 (in addition to $ℓn(g)$), or

Table 2. Formulas for obtaining expected values.

	Degree*		Eigenvector		Closeness		Betweenness	
	b	p	b	p	b	p	b	p
Intercept	−0.00063048		0.50379		1.74651		0.14623	
g			0.00059528	<0.0001	0.00458	0.009	0.00030033	<0.0001
density	1.00082	<0.0001	0.01969	<0.0001	0.28678	0.004	−0.02882	<0.0001
$ℓn(g)$			−0.10348	<0.0001	−0.36897	0.0003	−0.03423	<0.0001
R^2	0.9999		0.9635		0.1789		0.7116	

As an example of using these equations, a network with $g = 30$ actors and 25% density would yield expected values of betweenness centralities of 0.03 (read the table above within a column):

Expected Normed Betweenness Centrality

$$= 0.14623 + 0.00030033(g) - 0.02882(\text{density}) - 0.03423(ℓn(g))$$
$$= 0.14623 + 0.00030033(30) - 0.02882(0.25) - 0.03423(3.4012)$$
$$= 0.031612$$

The network's actual (normed) betweenness centralities may be compared to that expected base. Say there were 10 actors with betweenness centralities of 162.4 (normed betweenness of 0.4), 10 actors with betweenness centralities of 81.2 (normed at 0.2), and 10 actors with betweenness centralities of 0 (normed at 0.0). A X^2 may be calculated:

$$X^2 = 10\left[\frac{(0.4 - 0.03)^2}{0.03}\right] + 10\left[\frac{(0.2 - 0.03)^2}{0.03}\right] + 10\left[\frac{(0.0 - 0.03)^2}{0.03}\right]$$
$$= 10(4.563) + 10(0.963) + 10(0.030) = 45.63 + 9.63 + 0.30 = 55.56$$

On $(g - 1) = 29$ degrees of freedom, the critical value of chi-square is $\chi^2 = 42.56$, which the observed X^2 value exceeds; alternatively, the observed chi-square yields a probability value of 0.002. That is, for this hypothetical scenario, the set of 30 actors' betweenness centralities are significantly different from the values of betweenness that would be expected for a network with a random distribution of ties among 30 actors with 25% ties present. *Given the analytical derivation for expected degree, this prediction equation would be simply of the form: Expected Normed Degree Centrality = 1.0 (density). Expected centrality values may be calculating using the spreadsheet available from the authors or the SAS code in the Appendix.

the interaction between the number of actors and ties, but these terms were not significant. There was one outlier observation: for a normed closeness of 1.0, the standardized residual was very large, 3.0, but we did not delete that observation, because statistically speaking these random samples were as likely as any other random samples, and to purify the results in this manner seemed arbitrary and potentially misleading. Next, given that closeness centralities reflect distances, in a manner not true of degree, eigenvectors, or even betweenness, we sought alternatively to model the raw, non-normed closeness centralities (on the same network data) and were more successful: the regression equation,

$$0.279 + 0.00084g - 0.0075\text{density} - 0.0776\ell n(g),$$

resulted in an $R^2 = 0.562$. The coefficient for density was not significant, and the reduced model, $0.275 + 0.00084g - 0.0776\ell n(g)$, fit nearly as well, $R^2 = 0.559$, certainly better than the R^2 for the normed closeness values. Still, it seems reasonable to conclude that predicting expected values for closeness needs further study.

The formulas in **Table 2** may be used to calculate the values for degree, betweenness, and eigenvector centrality indices on a network with random ties (and more tentatively for closeness centralities). A network with no apparent inter-connected structure can serve as a useful benchmark, against which a real network should show significantly greater structure. **Table 2** contains an illustration of calculating the centrality indices for actors in a random network of 30 actors with 25% density. The expected values are then compared to the real centralities for the actors in such a network, and a chi-square is calculated to determine whether the observed centrality scores simply fall within the realm of random values, or do they exceed the lower and upper bounds of the random value to express significant network structures.

Table 3 illustrates still another use of these random network benchmark statistics. The table shows how to calculate confidence intervals for the centrality indices. Having obtained a confidence interval, of say [2.3, 13.9], a network researcher may compare the real network data centrality vector, say [1, 2, 3, 3, …, 12, 13, 15, 24, 27]. In this scenario, the observed centralities of values from 3 to 13 are contained in the confidence interval, indicating the actors whose centralities are not significantly different from those that would be expected if the network ties were simply random. The values 1 and 2 and 15 or greater fall outside the confidence interval, indicating these actors' patterns are distinct from random variability on network inter-connections. That is, for some reason the network researcher would investigate, the actors with only 1 or 2 or 15 or more ties are different from the others. The actors with centralities between 3 and 13 may still be interesting for other reasons, such as their forming cliques or equivalence blocks, but their volume of centralities are not particularly distinctive from networks of random collections of ties. Also note that the network researcher would conclude that overall, the network pattern was significantly different from a random network (just not the inner set of actors with values 3 to 13). A random

Table 3. Formulas for obtaining confidence intervals. Step 1: Calculate Expected Value, EV (from the appropriate equation in Table 2); Step 2: Calculate Expected Standard Deviation, ESD (in the following table); Step 3: form 95% confidence interval as $EV \pm t_{(g-1),\alpha/2}(ESD)(1/(g-1))$.

	Degree		Eigenvector		Closeness		Betweenness	
	b	p	b	p	b	p	b	p
Intercept	0.26271		0.27074		2.35759		0.18596	
g	0.00038748	<0.0001	0.00054664	<0.0001	0.00473	0.11	0.00050449	<0.0001
density			−0.06384	<0.0001	−0.67384	<0.0001	−0.03206	<0.0001
$\ell n(g)$	−0.05663	<0.0001	−0.06160	<0.0001	−0.53919	0.002	−0.04801	<0.0001
R^2	0.8511		0.7338		0.2472		0.6403	

To continue with the example of a network with $g = 30$ actors and 25% density, and an expected value for betweenness centralities of 0.03 from Table 2, the expected value for the standard deviation on the betweenness centralities would be:

$$S.D. = 0.18596 + 0.00050449(g) - 0.03206(\text{density}) - 0.04801(\ell n(g))$$
$$= 0.18596 + 0.00050449(30) - 0.03206(0.25) - 0.04801(3.4012)$$
$$= 0.29788 = 0.3$$

A 95% confidence interval is then calculated:

$$0.03 \pm t_{(29),.025}(0.3)(1/29) = 0.03 \pm 2.36(0.3)(1/29)$$
$$= 0.03 \pm 0.00244 = [0.027559, 0.32441]$$

If the observed normed betweenness centralities were once again, ten 0.4's, ten 0.2's, and ten 0.0's, we could conclude that the actors with 0.4's had significantly greater betweenness scores, and the actors with 0.0's had significantly lower betweenness centralities, than would be expected for a network with a random distribution of ties among 30 actors with 25% ties present. The actors with betweenness centralities of 0.2 had scores within the bounds of what would be expected in a random network. Confidence intervals may be calculating using the spreadsheet available from the authors or the SAS code in the Appendix.

network would contain all of its centrality values within the span of the random-derived confidence interval. A real network will very likely contain many of its centralities therein, but the difference is that some centralities will fall outside the bounds of the confidence interval.

The different states of the actors' centralities may be correlated with other information. Actors who have significantly lower or higher centralities may be of one political party affiliation, gender, ethnicity, or attitude compared to actors whose centralities fall in the span of random values. That is, these inferential tests allow networks researchers to draw conclusions beyond simply which actors are significantly different, but in conjunction with other explanatory variables, the distinctions and variability may be explained.

This investigation considered the usefulness of random networks, whereby the lack of structure provides a beneficial baseline to determine a stochastic likelihood of substantial structure over random structure. Study 2 used wholly random networks of varying size and densities to derive expected values, with a combination of analytical and empirical derivations. Illustrations of the calculation of a chi-squared statistic on the expected values, and confidence intervals using the expected values and their standard deviations, were also provided.

For convenience, software for calculating expected values of centralities as in

Table 2, or confidence intervals as in Table 3, is available from the authors in the form of Excel spreadsheets. In addition, the Appendix contains the SAS code for obtaining both expected values and confidence intervals.

The establishment of expected values, the chi-square tests, the z-score follow-up tests, and the confidence intervals are all important contributions to continue building on the inferential arm in the social network analysis literature. For both the network scholar and for the scholar's intended audience, tests of hypotheses enable conclusions about what effects in the network are "real" in a manner offered with greater statistical confidence than the presentation of merely descriptive statistics.

5. Discussion

This research considered four key centrality indices: degree, closeness, betweenness, and eigenvectors. Study 1 began with forms of stylized networks expected to exemplify conditions under which specific particular sensitivities of the four centralities should be clearest. Depending on the nature and content of the relational ties, some centrality indices seem more applicable or meaningful than others. As noise was added to the network ties, any distinctiveness was erased to the point that the average centrality indices converged across the network structures. This observation was suggestive, thus Study 2 focused on networks comprised entirely of random ties.

Study 2 focused on fully random networks. The examination of random ties allowed for the development of several comparative benchmarks. First, expected values were derived analytically for degree centralities and empirically for closeness, betweenness, and eigenvector centralities. The accuracies of the betweenness and eigenvector centrality estimators (and of course, degree) were at reasonably acceptable levels, but further research will be required to clarify the closeness estimators. Future research can also envelop non-binary and directed ties.

The expected values then enabled further tools for inferential tests of network structures, including first the chi-square test and its follow-up analyses for the micro-level examination, actor by actor to establish whether a set of actor centralities exceeded random patterns. Next, standard deviations were derived, which allowed for the construction of confidence intervals, similarly for the purpose of testing and demonstrating whether a set of observed centralities fell within the realm of random parameters or were significantly different from random, thereby indicating their more systematic and substantial patterns and natures.

In any given real network, many centrality values will be near the expected values, but the network as a whole would not be considered random unless all values were near their expected values, with none being statistically different. Instead, when actors' centralities are significantly different from values expected in random networks, researchers can be confident that there is indeed something structurally interesting about those actors. Furthermore, those differences may

be investigated to learn whether the different classes of actors (significantly lower, significantly higher, or random levels of centrality indices) are correlated with or explained by independent variables of theoretical interest to the network scholar.

This research builds further on the literature for inferential methods for analyzing social network data. Many centrality indices were originally created as descriptive statistics, without accompanying statistical distributions to test the significance of their observed values. The statistics offered in this paper were derived using random networks that served conceptually as a comparison. Observed centrality indices may now be tested against those standards to test the hypothesis as to whether the apparent network structure is random, or the pattern of ties is connected in a more meaningful way. Descriptive statistics are certainly informative, however an inferential approach goes a step further in allowing hypothesis testing about network structure, in turn enabling conclusions based less on subjective judgment and more on stronger grounds of statistical confidence.

We believe these techniques are easily implemented (see the **Appendix**). We hope they lend complementary insight to understanding actors in social network data.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Bonacich, P. (1987) Power and Centrality: A Family of Measures. *American Journal of Sociology*, **92**, 1170-1182. https://doi.org/10.1086/228631

[2] Mizruchi, M.S. and Potts, B.B. (1998) Centrality and Power Revisited: Actor Success in Group Decision Making. *Social Networks*, **20**, 353-387. https://doi.org/10.1016/S0378-8733(98)00009-4

[3] Freeman, L. (1978/1979) Centrality in Networks: Conceptual Clarification. *Social Networks*, **1**, 215-239.   https://doi.org/10.1016/0378-8733(78)90021-7

[4] Bonacich, P. (1972) Factoring and Weighting Approaches to Status Scores and Clique Identification. *Journal of Mathematical Sociology*, **2**, 113-120. https://doi.org/10.1080/0022250X.1972.9989806

[5] Knoke, D. and Yang, S. (2007) Social Network Analysis. 2nd Edition, Sage, Los Angeles, CA.

[6] Wasserman, S. and Faust, K. (1994) Social Network Analysis: Methods and Applications. Cambridge University Press, New York. https://doi.org/10.1017/CBO9780511815478

[7] Bolland, J.M. (1988) Sorting out Centrality: An Analysis of the Performance of Four Centrality Models in Real and Simulated Networks. *Social Networks*, **10**, 233-253. https://doi.org/10.1016/0378-8733(88)90014-7

[8] Lu, L., Chen, D., Ren, X., Zhang, Q., Zhang, Y. and Zhou, T. (2016) Vital Nodes Identification in Complex Networks. *Physics Reports*, **650**, 1-63. https://doi.org/10.1016/j.physrep.2016.06.007

[9] Opsahl, T., Agneessens, F. and Skvoretz, J. (2010) Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths. *Social Networks*, **32**, 245-251. https://doi.org/10.1016/j.socnet.2010.03.006

[10] Rothenberg, R.B., Potterat, J.J., Woodhouse, D.E., Darrow, W.W., Muth, S.Q. and Klovdahl, A.S. (1995) Choosing a Centrality Measure: Epidemiologic Correlates in the Colorado Springs Study of Social Networks. *Social Networks*, **17**, 273-297. https://doi.org/10.1016/0378-8733(95)00267-R

[11] Brandes, U., Borgatti, S.P. and Freeman, L.C. (2016) Maintaining the Duality of Closeness and Betweenness Centrality. *Social Networks*, **44**, 153-159. https://doi.org/10.1016/j.socnet.2015.08.003

[12] Kim, J., Ahn, H. and Kim, K. (2015) Performance Analysis of an Estimated Closeness Centrality Ranking Algorithm in Large-Scale Workflow-Supported Social Networks. *Journal of Internet Computing and Services*, **16**, 71-77. https://doi.org/10.7472/jksii.2015.16.3.71

[13] Smith, J.A. and Moody, J. (2013) Structural Effects of Network Sampling Coverage I: Nodes Missing at Random. *Social Networks*, **35**, 652-668. https://doi.org/10.1016/j.socnet.2013.09.003

[14] Stephenson, K. and Zelen, M. (1989) Rethinking Centrality: Methods and Examples. *Social Networks*, **11**, 1-37. https://doi.org/10.1016/0378-8733(89)90016-6

[15] Costenbader, E. and Valente, T.W. (2003) The Stability of Centrality Measures When Networks are Sampled. *Social Networks*, **25**, 283-307. https://doi.org/10.1016/S0378-8733(03)00012-1

[16] Riondata, M. and Kornaropoulos, E.M. (2016) Fast Approximation of Betweenness Centrality through Sampling. *Data Mining and Knowledge Discovery*, **30**, 438-475. https://doi.org/10.1007/s10618-015-0423-0

[17] Venkataramani, V., Zhou, L., Wang, M., Liao, H. and Shi, J. (2016) Social Networks and Employee Voice: The Influence of Team Members' and Team Leaders' Social Network Positions on Employee Voice. *Organizational Behavior and Human Decision Processes*, **132**, 37-46. https://doi.org/10.1016/j.obhdp.2015.12.001

[18] Xiao, M., Wu, J. and Huang, L. (2014) Community-Aware Opportunistic Routing in Mobile Social Networks. *IEEE Transactions on Computers*, **63**, 1682-1695. https://doi.org/10.1109/TC.2013.55

[19] Zemljič, B. and Hlebec, V. (2005) Reliability of Measures of Centrality and Prominence. *Social Networks*, **27**, 73-88. https://doi.org/10.1016/j.socnet.2004.11.010

[20] Katz, L. (1953) A New Status Index Derived from Sociometric Analysis. *Psychometrika*, **18**, 39-43. https://doi.org/10.1007/BF02289026

[21] Bonacich, P. (2007) Some Unique Properties of Eigenvector Centrality. *Social Networks*, **29**, 555-564. https://doi.org/10.1016/j.socnet.2007.04.002

[22] Bonacih, P. and Lloyd, P. (2015) Eigenvector Centrality and Structural Zeroes and Ones: When Is a Neighbor Not a Neighbor? *Social Networks*, **43**, 86-90.

[23] Borgatti, S., Carley, K.M. and Krackhardt, D. (2006) On the Robustness of Centrality Measures under Conditions of Imperfect Data. *Social Networks*, **28**, 124-136. https://doi.org/10.1016/j.socnet.2005.05.001

[24] Li, X., Liu, Y., Jiang, Y. and Liu, X. (2016) Identifying Social Influence in Complex Networks: A Novel Conductance Eigenvector Centrality Model. *Neurocomputing*, **210**, 141-154. https://doi.org/10.1016/j.neucom.2015.11.123

[25] Freeman, L.C., Borgatti, S.P. and White, D.R. (1991) Centrality in Valued Graphs: A Measure of Betweenness Based on Network Flow. *Social Networks*, **13**, 141-154.

https://doi.org/10.1016/0378-8733(91)90017-N

[26] White, D.R. and Borgatti, S.P. (1994) Betweenness Centrality Measures for Directed Graphs. *Social Networks*, **16**, 335-346.
https://doi.org/10.1016/0378-8733(94)90015-9

[27] Borgatti, S. and Everett, M. (1999) Models of Core/Periphery Structures. *Social Networks*, **21**, 375-395. https://doi.org/10.1016/S0378-8733(99)00019-2

[28] Watts, D.J. and Strogatz, S.H. (1998) Collective Dynamics of "Small-World" Networks. *Letters to Nature*, **393**, 440-442. https://doi.org/10.1038/30918

[29] Haberman, S. (1973) The Analysis of Residuals in Cross-Classified Tables. *Biometrics*, **29**, 205-220. https://doi.org/10.2307/2529686

[30] Hogg, R.V., Tanis, E. and Zimmerman, D. (2015) Probability and Statistical Inference. 9th Edition, Pearson, Upper Saddle River.

Appendix

SAS Code to Generate Expected Values and Confidence Intervals for Degree, Eigenvector, Closeness, and Betweenness Centralities

```
prociml;
g = 30; *<-- enter number of actors, g, here;
density = 0.25; *<-- enter density (or approximation) here;
degree = 1.0 × density;
print "expected degree" = "degree";
eigenv = 0.50379 + (0.00059528 × g) + (0.01969 × density) − (0.10348 × (log(g)));
print "expected eigenvector" = "eigenv";
closen = 10.74651 + (0.00458 × g) + (0.28678 × density) − (0.36897 × (log(g)));
print "expected closeness" = "closen";
between= 0.14623 + (0.00030033 × g) − (0.02882 × density) − (0.03423 × (log(g)));
print "expected betweenness" = "between";
sdd = 0.26271 + (0.00038748 × g) − (0.05663 × (log(g)));
sde = 0.27074 + (0.00054664 × g) − (0.06384 × density) − (0.0616 × (log(g)));
sdc = 20.35759 + (0.00473 × g) − (0.67384 × density) − (0.53919 × (log(g)));
sdb = 0.18596 + (0.00050449 × g) − (0.03206 × density) − (0.04801 × (log(g)));
lowcid = degree − (sdd × (1/(g−1)) × (tinv(0.975, (g−1))));
hicid = degree + (sdd × (1/(g−1)) × (tinv(0.975, (g−1))));
lowcie = eigenv − (sde × (1/(g−1)) × (tinv(0.975, (g−1))));
hicie = eigenv+ (sde × (1/(g−1)) × (tinv(0.975, (g−1))));
lowcic = closen − (sdc × (1/(g−1)) × (tinv(0.975, (g−1))));
hicic = closen + (sdc × (1/(g−1)) × (tinv(0.975, (g−1))));
lowcib = between − (sdb × (1/(g−1)) × (tinv(0.975, (g−1))));
hicib = between + (sdb × (1/(g−1)) × (tinv(0.975, (g−1))));
print "95% Confidence Interval for Degrees": " lowcid" to "hicid";
print "95% Confidence Interval for Eigenvector": "lowcie" to " hicie";
print "95% Confidence Interval for Closeness": "lowcic" to "hicic";
print "95% Confidence Interval for Betweenness": "lowcib" to "hicib";
quit; run;
```