

Network Topology Effects on Correlation between Centrality Measures

Ian McCulloh

*United States Military Academy,
Network Science Center, West Point, NY*

Centrality measures are often used to describe influential nodes in a network. When these measures are highly correlated they may be redundant and when they are uncorrelated they provide unique insight into the network. I propose a network simulation approach that creates networks with varying degrees of Erdos-Renyi randomness and Albert-Barabasi scale-freeness. Using this simulation approach I conduct 10 replications of a full factorial experimental design with varying levels of density and randomness versus scale-freeness. The effects of topology and density on the correlations of degree, betweenness, closeness, and eigenvector centrality are investigated. I find that not only does density and topology affect the correlation of centrality measures, but there exist many interaction effects as well. In general, networks with high Erdos-Renyi randomness tend to have higher levels of correlation between centrality measures than networks with Albert-Barabasi scale-freeness.

Author

Ian McCulloh, Ph.D. is an Assistant professor in the Department of Behavioral Sciences and Leadership at the United States Military Academy.

Acknowledgments: *I thank Kathleen Carley for the Organizational Risk Analyzer (ORA) which was used to simulate networks and calculate the centrality measures. Support for this research was provided by the Army Research Organization, Project No. 9FDATXR048. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Organization or the U.S. government.*

Correspondence: *Contact Ian McCulloh, Department of Behavioral Sciences and Leadership, at the United States Military Academy, 646 Swift Road, West Point, NY 10996; email: ian.mcculloh@usma.edu.*

INTRODUCTION

Centrality measures are often used in social network analysis to identify influential nodes. Among the most common centrality measures are degree, closeness, betweenness (Freeman, 1979), and eigenvector (Bonacich, 1972) centralities. Each of these centrality measures is based on a different concept of what makes a node influential or central to a network. When these measures are highly correlated, the differences in centrality measures are insignificant and redundant (Valente et al, 2008). When there is low correlation between centrality measures, they may offer unique insight into the network.

There have been a few studies that investigate the correlation between network centrality measures. Three studies investigated the correlation of measures under conditions of missing data (Bolland, 1998; Borgatti et al 2006; McCulloh, 2009). Other studies investigated the correlation between network centrality measures in applied network studies (Rothenberg, 1995; Faust, 1997; Valente and Forman, 1998). A more recent study investigated the correlation of centrality measures in empirical network data (Valente et al, 2008). This was an important study for revealing the effects of correlation on social network analysis. Valente also identifies the affect of density and directionality on the correlation of centrality measures. This paper expands upon their research by exploring how the underlying network topology effects these correlations. In using the concept of network topology, the findings can be more easily generalized to new network studies.

The idea of network topology is not clearly defined in the literature. The concept is used here only in the context of the degree distribution. Two network topologies are therefore compared: random and scale-free networks. Paul Erdős and Alfréd Rényi (1959) proposed the random network. In this model nodes are connected with some probability, p . Others have shown that the distribution of the degree will follow a binomial distribution

(Bolobos, 2003, McCulloh, 2009). Albert-Lazlo Barabasi and Reka Albert (1999) proposed an alternate model of network topology based on the theory of preferential attachment. The mechanism of preferential attachment was originally proposed by Herb Simon (1955) and suggests that when social groups are growing, new members will choose to connect to individuals with probability proportional to their prestige, often measured by degree (Wasserman and Faust, 1994). Simon, and later Barabasi and Albert, show that when preferential attachment exists in a growing network, the distribution of the degree will follow a power law. Under certain conditions this creates a network with a few central hub nodes. It has been shown that the effect of missing data on network measures is significantly affected by the degree topology of the network (Frantz et al, 2009). Networks that have strong tendencies for preferential attachment are more robust in their identification of top central nodes when there exists increasing levels of missing data.

In this paper, a simulation methodology is proposed that will generate networks that fall on a continuous spectrum between Erdős-Rényi random and Albert-Barabasi scale-free. Using this simulation methodology, networks are generated where the topology and density of the network are varied in a full-factorial, statistically designed experiment to explore the response surfaces of the correlation between network centrality measures.

In the next section the four centrality measures are briefly reviewed: degree, closeness, betweenness and eigenvector centrality. These measures were chosen for this study because they are the most common in the literature. The Methods section provides a description of how networks are simulated and an explanation of the virtual experiment. The results will be presented followed by discussion.

BACKGROUND

The *degree centrality* of node a is the number of other nodes directly connected to node a . The

degree centrality is a measure of direct influence in the network.

The *closeness centrality* of a node describes how close a node is to all others in the network. To calculate closeness, the geodesic path between all pairs of nodes in the network must be calculated. The closeness of node a then is the inverse of the average geodesic from node a to all other nodes in the network. A node high in closeness can disseminate information within the network more efficiently based on its position. It is therefore an ideal target for network diffusion.

The *betweenness centrality* of a node describes how frequently a node falls on the geodesic between other nodes. The betweenness of node a is the number of geodesic paths containing node a divided by the total number of geodesics in the network which is always $n(n-1)$, where n is the number of nodes in the network and we assume that there are no reflexive links. A node that is high in betweenness is influential in that it can connect otherwise disconnected subgroups within the network. High betweenness nodes serve as gate keepers or brokers of information and resources.

Eigenvector centrality is based on the concept that a node is influential to the extent that it connects to influential alters. Bonacich (1972) discovered that if the degree centrality of a node is modified by the degree centrality of its alters and this process is iterated many times, the result converges to the eigenvector of the symmetric adjacency matrix of the network.

The density of a network is the number of links in the network divided by the number of possible links which is always $n(n-1)$ when we assume that there are no reflexive links in a directed network. It can be shown that the maximum likelihood estimate of the parameter p in an Erdős-Rényi random network is the density. This relationship allows the density to be specified as an experimental factor in the statistically designed experiment to explore the

density and degree topology effects on centrality measure correlation.

METHODS

Networks can be simulated in a manner that allows them to achieve a hybrid topology between Erdős-Rényi randomness and Albert-Barabasi scale-freeness. Here, this is achieved by randomly assigning links between nodes as in a random graph for some percentage of the target order of the network. Then, the remaining nodes are added using preferential attachment. In this study, all networks consist of 100 nodes. For example, in the first virtual experiment, 20 percent of the nodes, or 20 nodes are randomly assigned links with a probability, $p = 0.1$ corresponding to the target density. The remaining 80 nodes are added to the network using preferential attachment. This results in a network that is 20 percent random and 80 percent scale-free. In the second virtual experiment 80 nodes are assigned random links, while 20 are added via preferential attachment. This results in a network that is 80 percent random and 20 percent scale-free. In this manner, networks can be simulated that have varying levels of randomness and density. These network variables are then varied in a statistically designed experiment.

A two factor, full factorial, statistically designed experiment is used to explore the response surface of the correlations between centrality measures as a function of the topology and density. The response variables are the 6 correlations between the four centrality measures. The independent factors are the network topology and the density. The high and low level of the topology is set at 80 percent and 20 percent random, respectively. The high and low level of the density is set at 0.3 and 0.1 respectively. A center point was also used, where the topology is set at 50 percent random and the density is 0.2, in order to estimate significant curvature in the response surface. 10 replications at each combination of independent factors are run to estimate standard error and make accurate inference. Table 1 displays the

design matrix for the series of virtual experiments. “Std Ord” refers to the standard order of the full factorial experimental design (Montgomery, 2005).

Table 1. Experimental Design

Std Ord	Reps	No. Nodes	Initially Random Nodes	Target Density
1	10	100	20	0.1
2	10	100	80	0.1
3	10	100	20	0.3
4	10	100	80	0.3
5	10	100	50	0.2

RESULTS

Topology and randomness/scale-freeness are all highly, statistically significant predictors of correlation between centrality measures for five of the six correlations. The randomness/scale-freeness factor does not appear to be significant in the degree centrality – eigenvector centrality correlation, however, there does appear to be an interaction effect. The half normal plot for the degree centrality – betweenness centrality correlation is presented in Figure 1.

A half normal plot is used to demonstrate the significance of independent factors on the correlation value. The blue squares represent the independent factors and the green triangles represent the degrees of freedom associated with the error term. The factors represented in the half normal plots are all highly significant with p-values less than 0.001. The notation “A” is used in the plot to represent the density factor. The notation “B” is used to represent the randomness/scale-freeness factor. The notation “AB” represents the interaction effect between the two other factors. An interaction implies that one factor has a different effect on the response, depending on the level of the other factor.

The half normal plots for the other correlations between centrality measures are similar to the

Figure 1. HNP Correlation Degree-Betweenness

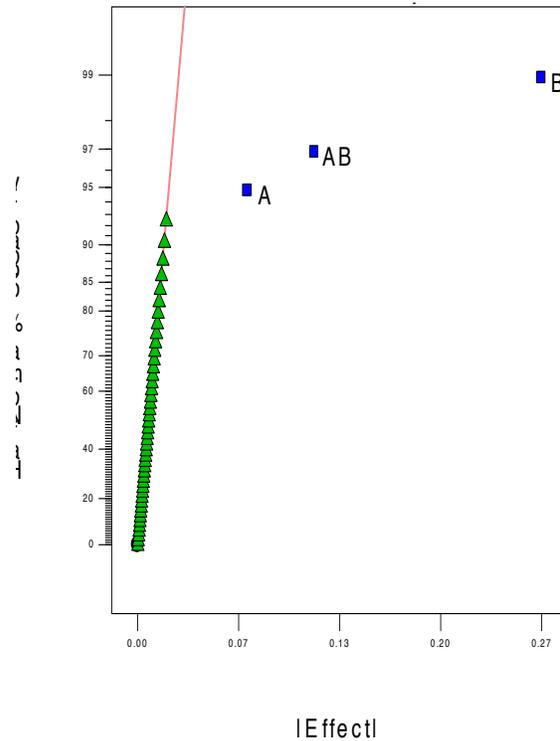


Figure 1. HNP (Half Normal Plot) is a graphical tool that uses ordered estimated effects to help assess which factors are significant and which are not.

degree – betweenness correlation; however, there are no interaction terms in the correlations involving closeness centrality. In addition, both independent factors (density and randomness) are significant in all correlations except for the degree-eigenvector centrality. For this correlation, the topology is only present in the interaction term.

It is important to also consider the degree to which topology and density affect the correlation between measures. Table 2 displays the coefficient of determination for the six response surfaces corresponding to the six correlations between centrality measures. The coefficient of determination quantifies how much of the variability in the correlation can be explained by the factor’s topology, density, and

interaction. It can be seen in Table 2 that the coefficients of determination are all very high, indicating that topology and density are perhaps the most significant factors affecting the correlation between centrality measures.

Table 2. Coefficients of Determination

Correlation	Coefficient of Determination
Degree-Betweenness	0.9474
Degree-Closeness	0.9178
Degree-Eigenvector	0.7812
Betweenness-Closeness	0.8526
Betweenness- Eigenvector	0.9221
Closeness-Eigenvector	0.9131

The actual response surfaces for the correlation between centrality measures are presented in Table 3. These equations can be used to calculate the expected correlation between centrality measures in a 100 node network, given a density and Erdos-Renyi randomness. The sign and magnitude of the coefficients in the

equation also provide insight into the effects of density and topology on the correlations. For example, the intercept term in the Degree-Eigenvector correlation is 0.95, which is an extremely high correlation. The coefficients for randomness and the interaction effect are small. This suggests that the density of the network has the largest effect on correlation between these measures. Because the coefficient of the density term is positive, the correlation between these measures will increase as the density increases.

A counter example of the Degree-Eigenvector correlation is found in the Degree-Closeness correlation. The intercept term for this response surface is low at 0.20. The coefficient for the randomness term is relatively high and negatively signed. This means that high randomness will lead to a more negative correlation between degree and closeness. In addition, an increase in the density of the network will contribute to negative correlations between the degree and closeness centrality measures.

Table 3. Response Surface Equations for the Correlation between Centrality Measures

Correlation	Final Equation in Terms of Actual Factors
Degree-Betweenness	$= 0.87 - 1.35d + 0.056r + 1.97dr$
Degree-Closeness	$= 0.21 - 0.55d - 0.74r$
Degree-Eigenvector	$= 0.95 + 0.12d + 0.01r - 0.08dr$
Betweenness-Closeness	$= -0.09 - 0.44d - 0.43r$
Betweenness- Eigenvector	$= 0.77 - 1.07d + 0.11r + 1.72dr$
Closeness-Eigenvector	$= 0.28 - 0.64d + 0.79r$

* d is the density of the network, and r is the Erdos-Renyi randomness of the network.

The relationships between the density, the randomness and the interaction effects can be illustrated with an interaction plot found in Figure 2. Figure 2 shows the interaction plot for the Degree – Betweenness centrality interaction. This figure plots the correlation of the two respective measures along the y-axis and the

randomness along the x-axis. There are two lines drawn in the plot corresponding to the high and low levels of network density. The points aligned with the middle of the x-axis are the center points in the experimental design. A significant interaction effect will cause the two lines to have different slopes, showing a different effect dependent upon the other factor.

There is a similar interaction for all of the other correlations with the exception of those involving closeness centrality. Since there is no interaction effect for closeness centrality, the two lines representing the different levels of density are parallel.

The correlations involving betweenness centrality reveal that increased density magnifies the effects of network topology on the correlation. As the density of the network and the randomness of the network increases, the betweenness centrality measure becomes more correlated with degree and eigenvector centrality. Recall that degree and eigenvector centrality are highly correlated for all networks in the study, so a measure correlated with degree will also be correlated with eigenvector centrality.

Figure 2. Interaction Plot Degree-Betweenness

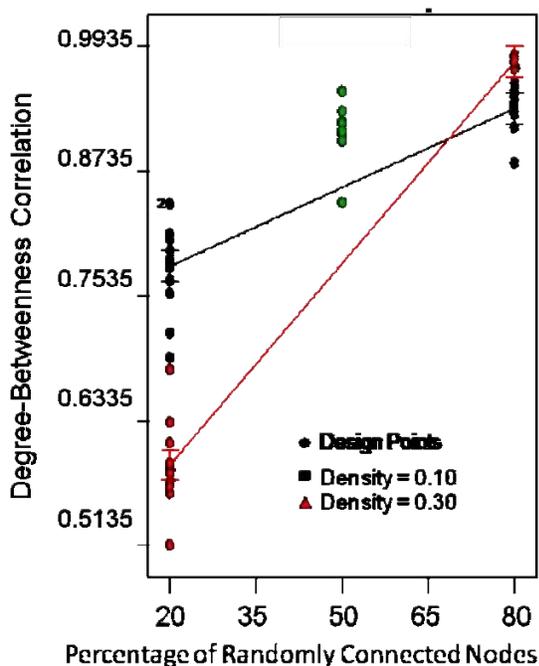


Figure 2. The X-axis is the percentage of nodes connected as an Erdős-Rényi Random graph (not preferential attachment). The Y-axis is the correlation between degree and betweenness centrality. The two lines correspond to the density. The line with greater slope has larger density, indicating that as density increases, topology has a greater affect on the correlation of centrality measures.

CONCLUSIONS

The nature of centrality measures in social network analysis is an understudied topic. The few papers investigating this area suggest that an improved understanding of the correlations between network measures are necessary to identify redundant measures versus measures that provide unique insight, as well as to understand the robustness of measures to missing data.

This paper shows that both network density and topology affect the correlation between centrality measures in the network. The correlations between measures of degree, betweenness, and eigenvector centrality exhibit interaction effects between the density and topology. Most compelling is that these two variables explain a vast majority of the variability in the correlation between measures. Finally, simulation provides a controlled environment for virtual experiments to explore network correlation studies, including a novel approach for modeling the scale-freeness of a network.

Only a limited number of measures were investigated in this paper. Future studies may include a wider range of network measures. There are many social network analysis software packages that include these network measures. The Organizational Risk Analyzer (ORA) available from Carnegie Mellon University provides most of these measures as well as a feature to generate stylized networks with varying levels of randomness or scale-freeness. In addition, this work only investigates networks with 100 nodes. It is possible that the order of the network will also affect the correlation of these measures. Based on the relationship between the order of the network and the density, it is unlikely that this would form a linear response surface. Hopefully, future researchers will consider other factors that might contribute significantly to the correlation between network measures.

The effect of topology on centrality measure correlation highlights its importance as a network property. Future investigations of network robustness to missing nodes, links, or attributes should consider network topology. In addition, simulation studies are likely to provide the most fruitful insight into network robustness in a similar approach as used by Borgatti and colleagues (2006). Their study of network robustness did not consider the topology of the network and was focused on Erdos-Renyi random networks. This study demonstrates the need for their important research to be replicated considering the affects of topology on the network.

In light of much discussion at academic conferences in regards to scale-free networks and preferential attachment, it is important to point out that this paper does not address how “real-world” networks evolve or in what ways networks exhibit randomness or scale-freeness. Many have commented on the issue of network topology in the literature and at academic conferences (Alderson, 2008; Doyle et al, 2005; Barabasi, 2008). The issue brought to light in this research is that network topology considerations will significantly affect the correlation between centrality measures. They are also likely to affect the outcomes of robustness studies. Therefore, future research should take network topology into consideration when presenting findings.

The correlation between network measures appears to be more correlated in random networks than in scale-free networks where hub nodes exist. This affect seems to be magnified by the density of the network when network density is high. This suggests that network measures may be redundant for Erdos-Renyi random networks, but that in other networks they may provide valuable, unique insight. Future research involving simulated networks should consider the impact of topology on their findings.

REFERENCES

- Alderson, D.L. (2008). Catching the “Network Science” Bug: Insight and Opportunity for the Operations Researcher. *Operations Research*, 56(5), 1047-1065
- Barabasi, A.L. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509-512.
- Barabasi, A.L. (2008). Keynote presentation at the 3rd US Military Academy Network Science Workshop, West Point, NY, 15 October 2008.
- Bolland, J.M. (1998). Sorting Out Centrality: An analysis of the Performance of Four Centrality Models in real and simulated networks, *Social Networks*, 10, 233-253.
- Bollobas, B. (1998). *Modern Graph Theory*. Springer-Verlag, New York.
- Bonacich, P. (1972). Technique for Analyzing Overlapping Memberships, *Sociological Methodology*, 4, 176-185.
- Borgatti, S.P., Carley, K.M., & Krackhardt, D. (2006). On the Robustness of Centrality Measures Under Conditions of Imperfect Data, *Social Networks*, 28, 124-136.
- Carley, K. M. (2009). *ORA: Organizational Risk Analyzer v.1.9.5*. [Network Analysis Software]. Pittsburgh: Carnegie Mellon University.
- Doyle, J., Alderson, D., Li, L., Low, S., Roughan, M., Shalunov, S., Tanaka, R., & Willinger, W. (2005). The “robust yet fragile” nature of the Internet. In: *Proceedings of the National Academies of Science*, 102(41), 14497–14502.
- Erdos, P, & Renyi, A. (1959). On random graphs. *Publicationes Mathematicae Debrecen*, 6, 290-297.
- Faust, K. (1997). Centrality in affiliation networks. *Social Networks*, 19, 157-191.
- Frantz, T., McCulloh, I., & Carley, K.M. (2009). Estimating the Reliability of Top-Actor Identification. *Network Science*, 2, to appear.
- Freeman, L. (1979). Centrality in social networks: I, conceptual clarification. *Social Networks*, 1 (1979), 215-239.

- Freeman, L. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40 (1977), 35-41.
- McCulloh, I. (2009). *Detecting Changes in a Dynamic Social Network*. Ph.D. Dissertation, Carnegie Mellon University.
- Montgomery, D.C. (2005). *Design and Analysis of Experiments*. 5th Ed. Wiley, Hoboken, NJ.
- Rothenberg, R. Potterat, J.J., Woodhouse, D.E., Darrow, W.W., Muth, S.Q., & Klovdahl, A.S. (1995). Choosing a centrality measure: Epidemiologic correlates in the Colorado Springs study of social networks. *Social Networks*, 17,273-297.
- Simon, H. (1955). On a class of skew distribution functions. *Biometrika* 42, 435-440.
- Valente, T.W., & Foreman, R.K. (1998). Integration and radiality: Measuring the extent of an individual's connectedness and reachability in a network. *Social Networks*, 20, 89-109.
- Valente, T.W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How Correlated Are Network Centrality Measures? *Connections*, 28, 16-26.
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.