Social Networks 50 (2017) 46–54

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Correlations among centrality indices and a class of uniquely ranked graphs☆

David Schoch[a,*], Thomas W. Valente[b], Ulrik Brandes[a]

[a] Department of Computer & Information Science, University of Konstanz, Germany
[b] Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

ARTICLE INFO

Article history:
Available online 5 April 2017

Keywords:
Network centrality
Correlation
Neighborhood inclusion
Centrality indices
Threshold graphs

ABSTRACT

Various centrality indices have been proposed to capture different aspects of structural importance but relations among them are largely unexplained. The most common strategy appears to be the pairwise comparison of centrality indices via correlation. While correlation between centralities is often read as an inherent property of the indices, we argue that it is confounded by network structure in a systematic way. In fact, correlations may be even more indicative of network structure than of relationships between indices. This has substantial implications for the interpretation of centrality effects as it implies that competing explanations embodied in different indices cannot be separated from each other if the network structure is close to a certain generalization of star graphs.

1. Introduction

Many centrality indices have been proposed to date and the list is ever-expanding (Todeschini and Consonni, 2009; Lü et al., 2016). In addition to their application in empirical research, they are a frequent subject of methodological work aiming to provide a better understanding of what centrality indices measure and the theoretical foundations of the concept as a whole (Freeman, 1979; Sabidussi, 1966; Nieminen, 1974; Borgatti, 2005; Borgatti and Everett, 2006; Boldi and Vigna, 2014; Schoch and Brandes, 2016).

A frequently investigated question in this context deals with correlations among centrality indices (Bolland, 1988; Rothenberg et al., 1995; Lee, 2006; Valente et al., 2008; Batool and Niazi, 2014; Li et al., 2015; Lozares et al., 2015). The underlying assumption being that correlations are a consequence of the formal definition of indices and thus highlight differences in the conceptualization of centrality. High correlation "suggests considerable redundancy" (Bolland, 1988) and thus justifies, e.g., the use of a computationally less expensive index (Li et al., 2015). Weakly correlated indices on the other hand "indicate distinctive measures likely to be associated with different outcomes" (Valente et al., 2008). It is thus not surprising that a correlation analysis is often performed when new indices are introduced to illustrate their disparity from existing measures (Newman, 2005; Estrada et al., 2009; Chen et al., 2012; Benzi and Klymko, 2013).

Reported results, however, are often inconsistent with regard to the similarity of centrality indices. Bolland (1988) finds that closeness, degree and eigenvector centrality are substantially correlated, yet betweenness was comparatively uncorrelated with other indices. Rothenberg et al. (1995) use eight different indices on a network of HIV patients and determine that all measures are highly correlated, including degree and betweenness. Likewise, Lee (2006) observes a high correlation between degree and betweenness on a set of protein interaction networks. Based on a broader sample of networks of varying origin as well as a set of random graphs, Batool and Niazi (2014) find that, overall, closeness and eccentricity as well as degree and eigenvector centrality are highly correlated and that correlations with betweenness vary across networks.

These inconsistencies suggest that the role of the underlying network structure is far more important and profound than is accounted for. Indeed, structure appears to be of interest mostly when the stability of centrality indices in the face of missing data or sampled networks is investigated (Frantz et al., 2009; Borgatti et al., 2006; Costenbader and Valente, 2003; Kim and Jeong, 2007). Borgatti et al. (2006) show that indices behave similar in terms of

☆ We gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft under grant Br 2158/6-1 and the Graduate School of Decision Sciences at the University of Konstanz. A preliminary version of this research was presented at the Sunbelt XXXIV Social Networks Conference (St. Pete Beach, Florida, February 2014).

* Corresponding author.
E-mail addresses: david.schoch@uni-konstanz.de (D. Schoch), tvalente@usc.edu (T.W. Valente), ulrik.brandes@uni-konstanz.de (U. Brandes).

http://dx.doi.org/10.1016/j.socnet.2017.03.010

Table 1
Characteristics of the networks compiled from various studies (replication of Table 2 from Valente et al., 2008).

Study	Networks	Average size	Average density	Average outdegree	Symmetrized centralization	In-degree centralization	Out-degree centralization
1	3	64	0.06	2.61	0.15	0.19	0.12
2	25	68	0.03	1.62	0.11	0.20	0.05
3	11	76	0.03	1.94	0.16	0.29	0.06
4	9	83	0.05	3.56	0.15	0.28	0.02
5	9	82	0.50	39.17	0.30	0.15	0.49
6	1	71	0.32	22.15	0.31	0.30	0.38
7	1	72	0.20	14.19	0.18	0.24	0.28
8	1	60	0.09	5.23	0.10	0.10	0.10

change patterns and level of robustness when edges are added or deleted on simple random graphs. Frantz et al. (2009) perform a similar analysis on more complex network structures and show that correlation varies with the used network model. Costenbader and Valente (2003) use a set of 60 empirical networks and examine the stability of indices on sampled networks and conclude that stability varies among indices. Again, they do concede that stability varies across networks.

Although many of the above studies already point out that structural properties of networks such as density (Valente et al., 2008) or degree heterogeneity (Kim and Jeong, 2007) impact correlations among indices, the assumption that correlations mainly depend on formal aspects appears to prevail.

In contrast to previous work, we show that in some cases correlations among indices are in fact dominated by structural properties. These are not necessarily visible in common network statistics such as density or degree distribution, but related to the class of threshold graphs (Mahadev and Peled, 1995) which generalize star graphs while maintaining an indisputed centrality ranking. The relevance of this class is due to the completeness of the neighborhood-inclusion preorder which in turn is preserved by all common centrality indices (Schoch and Brandes, 2016).

We start with a re-examination of a broad correlation study of Valente et al. (2008) (henceforth referred to as the *Correlation Study*), where varying correlations among a large set of indices were observed. Subsequently, we introduce the concept of neighborhood-inclusion and argue that the completeness of the ranking defined by neighborhood-inclusion is indicative for high correlations among indices, even for dual measures such as betweenness and closeness (Brandes et al., 2016). We then illustrate this concept on data from the Correlation Study and thus provide an alternative explanation for previously observed correlations. We conclude with implications of these findings for empirical research.

2. Results of the correlation study

Valente et al. (2008) use 60 networks compiled from eight different studies in varying contexts (Valente, 1995; Coleman et al., 1966; Burt, 1987; Rogers and Kincaid, 1981; Valente et al., 1997). All studies were conducted in bounded communities by interviewing their members, asking about relationships with other members. A more detailed description of the data can be found in Valente et al. (2008). It is important to note that the studies greatly differ in network size, number of networks, type of questions asked, and the number of nominations allowed. This leads to a diverse set of networks with varying structural properties. Initially the networks are directed. The authors, however, also consider symmetrized versions of the networks to compare undirected centrality measures. Our main analysis will focus on these indices for undirected networks; results for the directed cases are given in this section only for verification and completeness. Table 1 shows basic summary statistics of the eight studies and 60 directed networks.

Twelve different centrality indices are used in the correlation analysis. They include (in/out/symmetric) degree, (in/out/symmetric) closeness (Sabidussi, 1966), (in/out/symmetric) betweenness (Freeman, 1977), eigenvector centrality (Bonacich, 1972) as well as integration, radiality (Valente and Foreman, 1998) and their identical symmetric version. Correlations are assessed with Pearson's coefficient. Table 2 summarizes the average correlation among the twelve indices across the 60 networks.

Note that the symmetrized versions of degree, betweenness, closeness and eigenvector centrality are, on average, highly correlated. The larger differences between in and out measures can mostly be attributed to the nomination scheme of the individual studies, i.e. individuals nominate more people than they are themselves nominated.

Table 2
Average correlation between centrality indices across 60 networks. Symmetrized indices in bold (replication of Table 3 from Valente et al., 2008).

	1	2	**3**	4	**5**	6	7	**8**	9	10	**11**	**12**
1 Indegree												
2 Outdegree	0.28											
3 **Degree**	0.81	0.73										
4 Betweeness	0.62	0.55	0.75									
5 **S-betweeness**	0.70	0.51	**0.83**	0.71								
6 Closeness-in	0.60	0.19	0.48	0.37	0.32							
7 Closeness-out	0.20	0.82	0.58	0.40	0.37	0.07						
8 **S-closeness**	0.45	0.63	**0.62**	0.37	**0.40**	0.50	0.64					
9 Integration	0.74	0.28	0.61	0.50	0.42	0.93	0.18	0.57				
10 Radiality	0.23	0.87	0.62	0.45	0.40	0.12	0.98	0.66	0.23			
11 **S-int/rad**	0.50	0.69	**0.69**	0.43	**0.47**	0.52	0.68	**0.99**	0.60	0.71		
12 **Eigenvector**	0.74	0.70	**0.91**	0.67	**0.69**	0.46	0.54	**0.57**	0.60	0.59	**0.65**	
Average	0.54	0.57	0.69	0.53	0.53	0.41	0.50	0.58	0.51	0.53	0.63	0.65
Standard deviation	0.22	0.23	0.13	0.14	0.17	0.25	0.28	0.17	0.23	0.28	0.16	0.12