ALTERNATIVES TO BETWEENNESS CENTRALITY: A MEASURE OF CORRELATION COEFFICIENT

Xiaojia He[1] and Natarajan Meghanathan[2]

[1]University of Georgia, GA, USA,
[2]Jackson State University, MS, USA
[2]natarajan.meghanathan@jsums.edu

ABSTRACT

In this paper, we measure and analyze the correlation of betweenness centrality (BWC) to five centrality measures, including eigenvector centrality (EVC), degree centrality (DEG), clustering coefficient centrality (CCC), farness centrality (FRC), and closeness centrality (CLC). We simulate the evolution of random networks and small-world networks to test the correlation between BWC and the five measures. Additionally, nine real-world networks are also involved in our present study to further examine the correlation. We find that DEG is highly correlated to BWC on most cases and can serve as alternative to computationally-expensive BWC. Moreover, EVC, CLC and FRC are also good candidates to replace BWC on random networks. Although it is not a perfect correlation for all the real-world networks, there still exists a relatively good correlation between BWC and other three measures (CLC, FRC and EVC) on some networks. Our findings in this paper can help us understand how BWC correlates to other centrality measures and when to decide a good alternative to BWC.

KEYWORDS

Betweenness Centrality, Random Networks, Small-World Networks, Real-World Networks, Correlation Coefficient

1. INTRODUCTION

Over the past decades, a large number of centrality measures have been introduced and developed to quantify the significance and importance of the nodes in various networks. Betweenness centrality (BWC) is one of the most widely used measures, first developed in the 1970s by Freeman [1] and Anthonisse [2], independently. BWC is a measure of the degree to which a node functions as the mediation node by calculating the fraction score of all shortest paths (geodesic) between other pairs of nodes that go through it. It is expected that the network would be disconnected if one or two nodes with high BWC were removed. Thus one can expect that a node with high BWC does not belong to one of the dense groups, but connects them. For the rest of the paper, the terms 'node' and 'vertex', 'link' and 'edge', 'network' and 'graph' are used interchangeably. They mean the same.

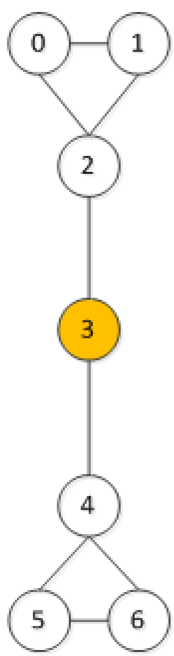
David C. Wyld et al. (Eds) : ICAITA, CDKP, CMC, SOFT, SAI - 2016
pp. 01– 10, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61301

BWC has been widely applied to a large number of complex network analyses. For instance, it has been proposed as an indicator of the "interdisciplinary" nature of scientific journals [3]. In general, BWC of the nodes in a network increases with connectivity as a power law with an exponent η [4]. Thus, it is known to be computationally time consuming to obtain exact BWC: O(nm) time for unweighted graphs and O($nm + n^2\log n$) time for weighted graphs, where n is the number of vertices and m is the number of edges in the network [5][6][14]. In this paper, we focus on analyzing the correlation between BWC and five well-known centrality measures, including eigenvector centrality (EVC), degree centrality (DEG), clustering coefficient centrality (CCC), farness centrality (FRC), and closeness centrality (CLC). Random network, small-world network, and several real-world networks are involved in this paper.

2. COMPUTATION OF BETWEENNESS CENTRALITY

The computation of BWC in this paper follows the algorithm by Brandes (2001) [5]. If the number of shortest paths between two nodes i and j that pass through node k as the intermediate node is denoted as g_{ij}^{k} and the total number of geodesic between the two nodes i and j is denoted as g_{ij}, then the BWC for node k is defined as

$$BWC(k) = \sum_i \sum_j \frac{g_{ijk}}{g_{ij}}(i \neq j \neq k)$$



Betweenness of node 2:
Node 3→ 0: 1
Node 3→ 1: 1
Node 4→ 0:1
Node 4→1:1
Node 5→0:1
Node 5→1:1
Node 6→0:1
Node 6→1:1
Betweenness Centrality of node 2:8

Betweenness of node 3:
Node 4→ 0: 1
Node 4→ 1: 1
Node 4→ 2:1
Node 5→0:1
Node 5→1:1
Node 5→2:1
Node 6→0:1
Node 6→1:1
Node 6→2:1
Betweenness Centrality of node 3:9

Betweenness of node 4:
Node 5→0:1
Node 5→1:1
Node 5→2:1
Node 5→3:1
Node 6→0:1
Node 6→1:1
Node 6→2:1
Node 6→3:1
Betweenness Centrality of node 4:8

Betweenness Centrality of node 0: 0
Betweenness Centrality of node 1: 0
Betweenness Centrality of node 5: 0
Betweenness Centrality of node 6: 0

Figure 1: Representative Example to Compute the Betweenness Centrality of the Vertices in a Network

The representative BWC calculation is illustrated in Figure 1. On the basis of the algorithm proposed by Brandes (2001) [5], breadth-first search is involved in the computation. It is clear that BWC is different to degree-based ranking as shown in Figure 1. Nodes 3 and 4 have highest degree in this present network; however, node 3 has highest BWC. Nodes 0, 1, 5, and 6 each has a degree of 2, but with a BWC of 0.

3. CORRELATION ANALYSIS

3.1. Analysis on Random Networks

Firstly, random networks were simulated to investigate all the six centrality measures including BWC, EVC, DEG, CCC, FRC, and CLC. In this section, networks with 100 nodes were simulated. Particularly, the probability of linkage between nodes is varied from 0.05 to 0.9 to evaluate above mentioned centrality measures. The probability of linkage is increased from 0.05 to 0.1 by 0.01; from 0.1 to 0.9 by 0.1. Representative random networks are shown in Figure 2 with a ranking factor of BWC. Correlation between BWC and other five measures, including DEG, EVC, CCC, FRC, and CLC, was then determined. Average correlation coefficient value was calculated based on 100 trials.

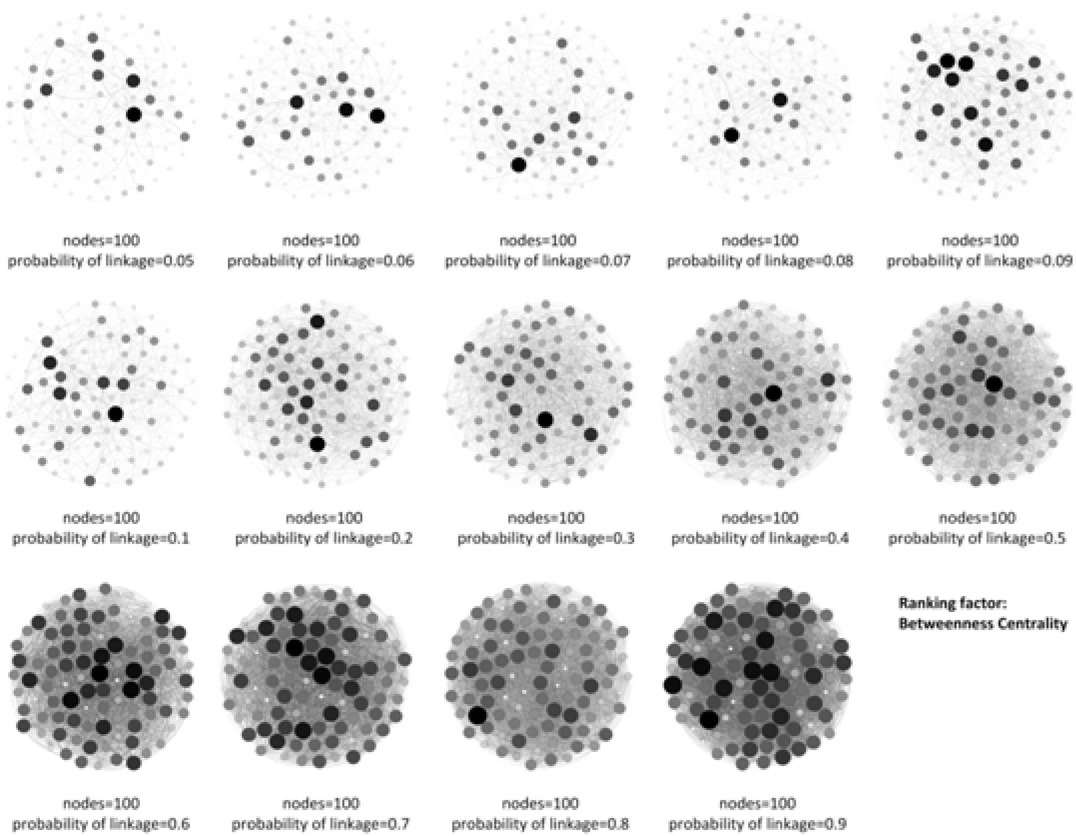


Figure 2: Simulation of Random Networks with Various Probability of Linkage Values
[Ranking Factor is Betweenness Centrality]

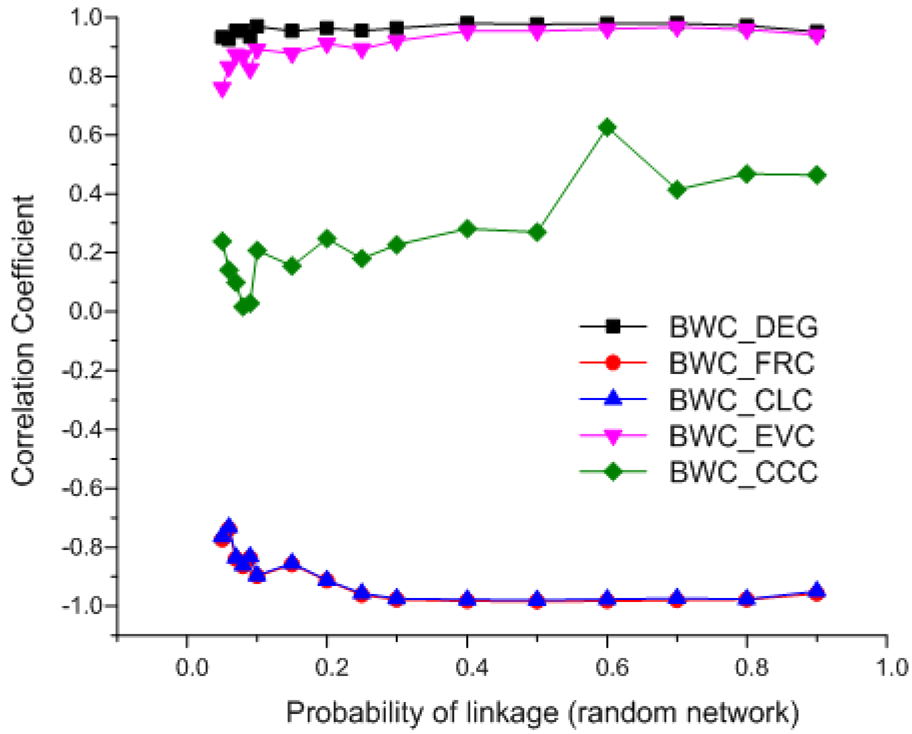


Figure 3: Correlation Coefficient between BWC and the other Five Centrality Measures: DEG, EVC, CCC, FRC and CLC on Random Networks with Various Probability of Linkage Values

As shown in Figure 3, BWC is highly correlated to all measures except CCC. Our data suggests a strong correlation between BWC and DEG, ranging from 0.9316 to 0.9513. The highest correlation of BWC to FRC, CLC, and EVC reaches -0.9576, -0.9495, and 0.94, respectively. The negative correlation indicates that an increase in one variable reliably predicts a decrease in the other one. A high value in negative correlation still suggests high correlation. It is pretty sure that we can select DEG, FRC, CLC, EVC as alternatives to BWC in random networks.

3.2. Analysis on Small-World Networks

We investigated on small-world networks evolved from regular network. Similar to random network simulation, 100 nodes with a k-regular value (initial number of links per node) of 10 are set for small-world network simulation. In this section, the probability of rewiring was varied from 0.01 to 0.09 with increment of 0.01; and from 0.1 to 0.9 with increment of 0.1. Representative small-world networks are shown in Figure 4 with a ranking factor of BWC. Correlation between BWC and the other five measures, including DEG, EVC, CCC, FRC, and CLC, was then calculated. Average correlation coefficient value was calculated based on 100 trials.

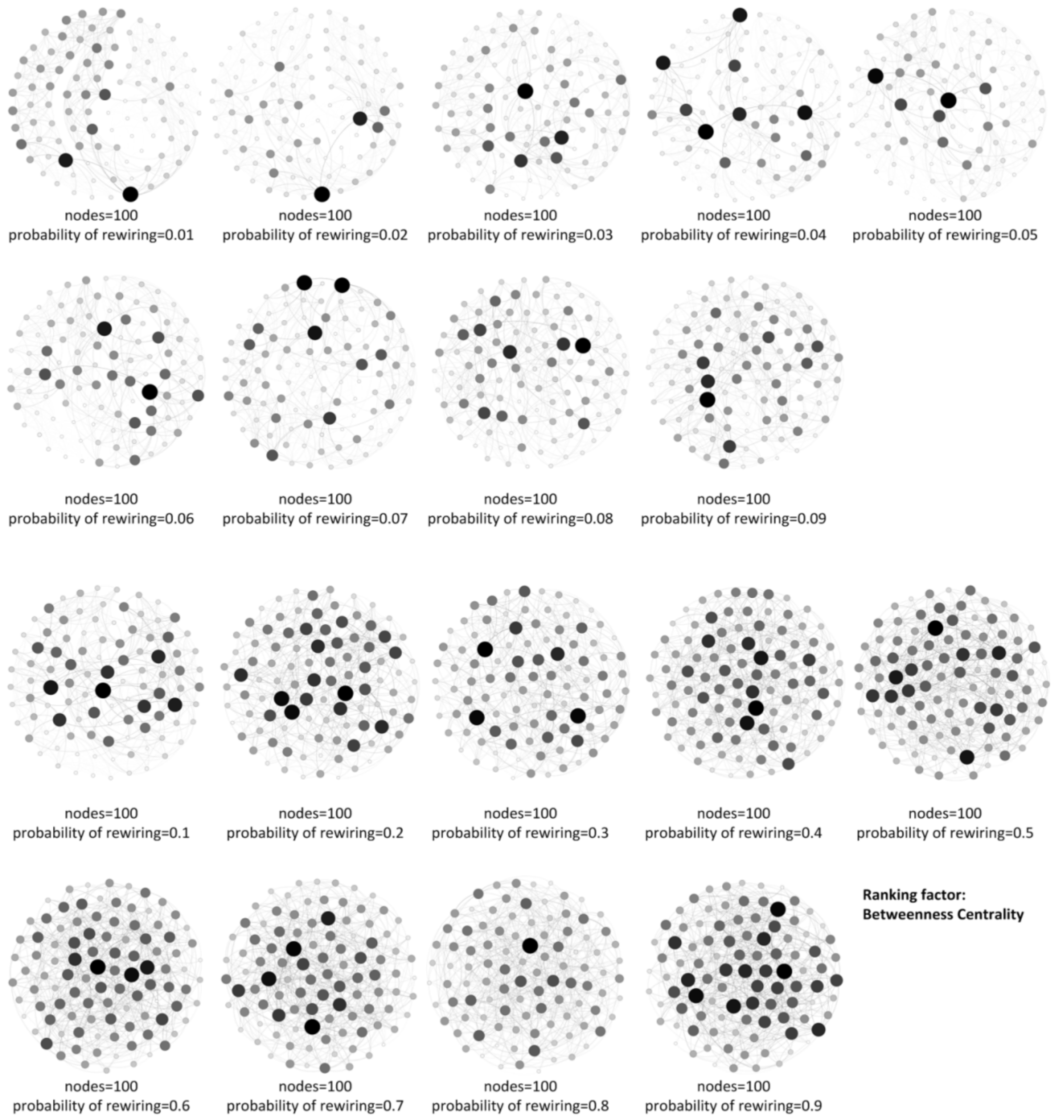


Figure 4: Simulation of Small-World Networks with Various Probability of Rewiring Values [Ranking Factor is Betweenness Centrality]

For small-world networks, there is a strong correlation between BWC and the other centrality metrics, except EVC, at a probability of rewiring lower than 0.2. The correlation coefficient was larger than 0.51 when the probability of rewiring reaches 0.2 for DEG, FRC, CLC, and CCC. The highest correlation coefficient of BWC to DEG, FRC, and CLC reaches to 0.5325, -0.7499, and -0.7348 at probability of rewiring of 0.08. The correlation between BWC and CCC decreases from 0.8131 to 0.0683 along with the increase of probability of rewiring.

In a previous work, a transformation between small-world network and random network was revealed [15]. It was found that simulated network from a regular network would be small-world network when the probability of rewiring is from 0.01 to 0.1; however, it changes to random network when the probability of rewiring is between 0.1 and 1.0. In this study, we also observed a clear turning point at probability of rewiring of 0.1 as shown in Figure 5. Overall, we could preferably use CCC as alternative to BWC at probability of rewiring lower than 0.07. At a critical probability of rewiring lower than 0.2, we still could use DEG, FRC, CLC, and CCC as alternatives.

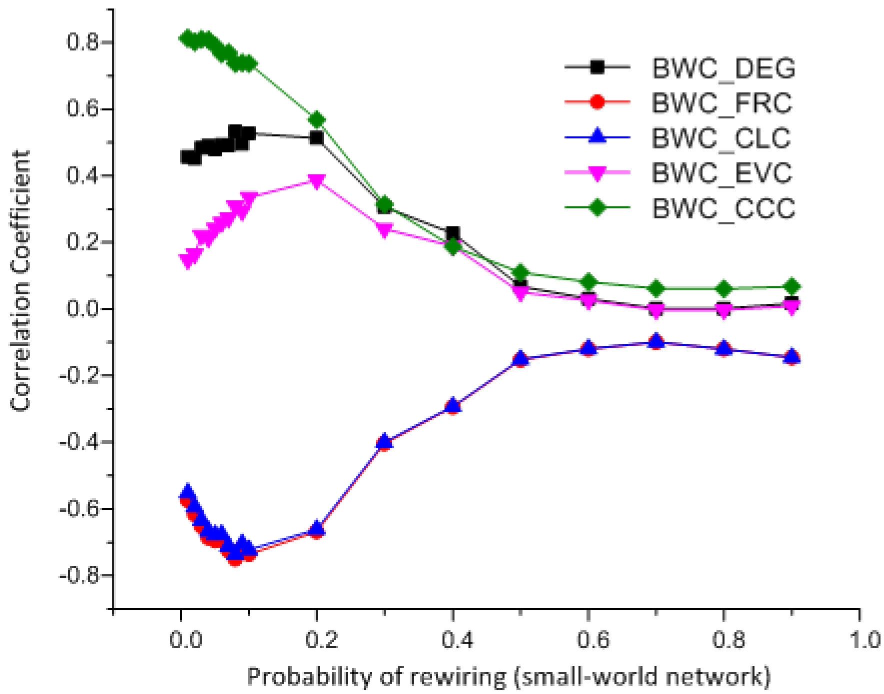


Figure 5: Correlation Coefficient between BWC and the other Five Centrality Measures, including DEG, EVC, CCC, FRC and CLC, on Small-World Networks with Various Probability of Rewiring Values

3.3. Analysis on Real-World Networks

In order to evaluate the feasibility of applying the above mentioned candidate centrality metrics to replace BWC practically, multiple real-world networks were also studied. Analysis on real-world networks is crucial to understanding how BWC relates to other measures in real world. In this study, nine real-world networks (see Figure 6) were analyzed. These are: Dolphins social network (Dolphins), Word adjacency network of common adjectives and nouns in the novel David Copperfield by Charles Dickens (WordAdj), Celegensmetabolic network representing the metabolic network of C. elegans (Celegm), Celegensneural network representing the neural network of C. elegans (Celegn), American football games network between Division IA colleges during regular season Fall 2000 (Football), Karate Social network of friendships between 34 members of a karate club at a US university in the 1970 (Karate), LesMis Coappearance network of characters in the novel Les Miserables (LesMis), the 1997 US Airports network (AirNet), and Political books network (BookNet). Average correlation between BWC and other five measures, including DEG, EVC, CCC, FRC, and CLC, was determined on 100 trials.

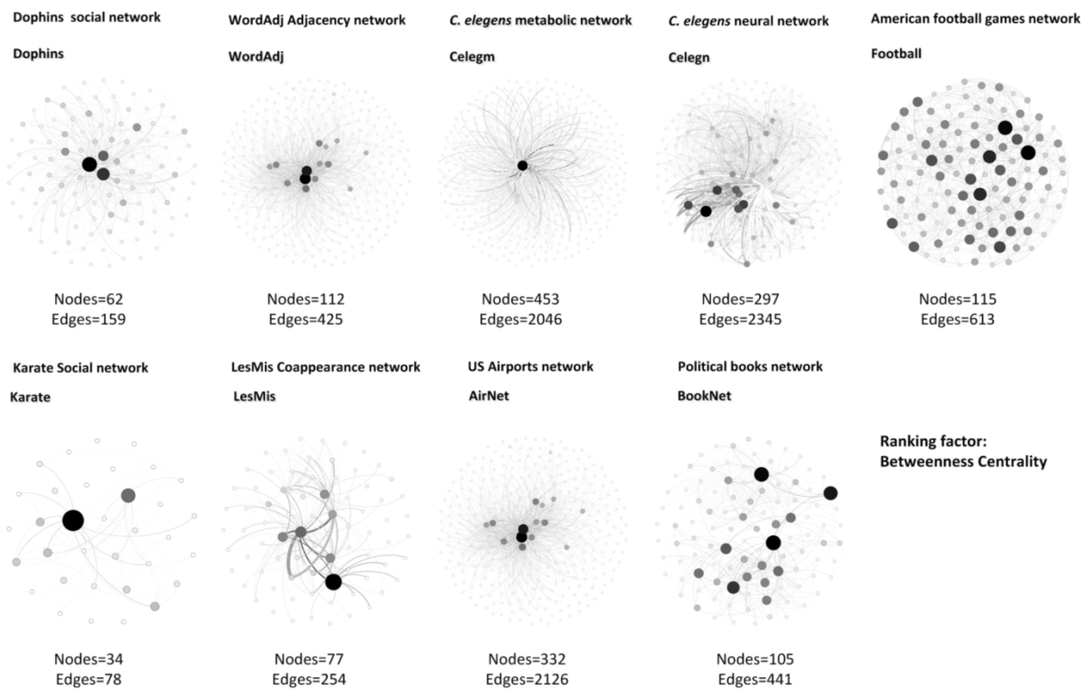


Figure 6: Distribution of the Nodes in Real-World Networks
[Ranking Factor: Betweenness Centrality]

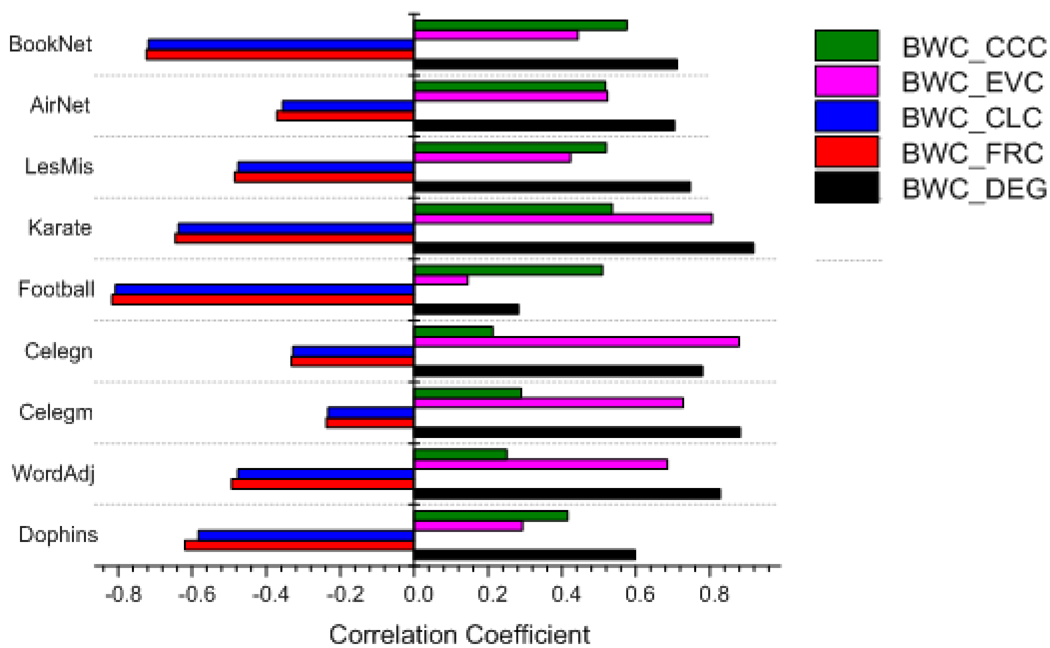


Figure 7: Correlation Coefficient between BWC and the other Five Centrality Measures, including DEG, EVC, CCC, FRC and CLC on Real-World Networks

Unlike the random and small-world networks, the correlation of BWC to CLC and FRC is relatively low with a correlation coefficient value less than 0.6 for five tested networks out of nine (WordAdj, Celegm, Celegn, LesMis, and AirNet). Similarly, correlation coefficient between BWC and EVC is also relatively low with a value less than 0.6 for five tested networks including Dolphins, Football, LesMis, AirNet, and BookNet. Particularly, the correlation coefficient between BWC and EVC on Football network only shows a value of 0.14. It is also similar to BWC and DEG with a correlation coefficient of 0.28 on Football network. Notably, the correlation coefficient between BWC and CCC is lower than 0.6 on all tested networks. It is noteworthy that BWC correlates well with DEG on all but Football network. On Football network, BWC has a high correlation with FRC and CLC

4. RELATED WORK

Recently, Meghanathan (2016) proposed a hybrid centrality metric (takes both the degree and the shortest paths into account) called the local clustering coefficient-based degree centrality (LCCDC) [10]. The local clustering coefficient (LCC) of a vertex is a measure of the probability that any two neighbors of the vertex are connected. If a vertex has a larger LCC value, then the neighbors of the vertex can directly communicate with each other rather than going through the particular vertex. If the neighbors of a vertex do not need go through the vertex for shortest path communication, then it is more likely that the rest of the vertices in the network would not need to go through the vertex for shortest path communication. If a vertex has a smaller LCC, then the neighbors of the vertex are more likely to go through the vertex for shortest path communication between themselves (as there is more likely not a direct edge between the two neighbors, because of the low LCC for the vertex). More specifically, if a vertex has a low LCC and a high degree, then several of the neighbors of the vertex (and as a result, several of the two-hop, three-hop neighbors and so on) are more likely to go through the vertex for shortest path communication. Such vertices are expected to have a higher BWC. The LCCDC metric captures such high BWC vertices (with a strongly positive correlation) and could be used to rank the vertices in a graph in lieu of the BWC. Since the strongly positive correlation between BWC and LCCDC has been already studied in [10], in this paper, we explore any of the other well-known centrality metrics exhibit a strong correlation with BWC.

There are some other algorithms proposed to further develop the application of BWC. For instance, the random-walk betweenness measure calculated for all vertices in a network in worst-case time $O((m+n)n^2)$ using matrix methods [8]. Others such as bounded-distance betweenness [9], distance-scaled betweenness [9], edge betweenness [11] and group betweenness [12] are also introduced. Alternatively, an approximation computation of BWC of a given vertex with an adaptive sampling technique is discussed in the paper by Bader et al (2007) [7]. Nevertheless, the computation cost of these betweenness measures is still high. It is more feasible if we could find another centrality measure with low computation cost that is highly correlated to BWC. It was shown that the BWC is related to the degree in social networks [13] and scale-free network [14]. However, there still lacks substantial support on the alternatives to BWC.

BWC measures the interrelationships among vertices. The results of our simulation studies suggest that BWC is highly correlated to DEG on most tested networks. Leydesdorff (2007) [3] also observed high correlation between BWC and DEG with a correlation coefficient value of 0.724 on Journal Citing Social Networks. Recently, Pozzi et al (2013) [16] observed a strong correlation of the centrality indices between unweighted BWC and DEG calculated on Planar

Maximally Filtered Graphs (PMFG) with a value of 0.97. There is also a moderate correlation between BWC and CLC papered with a value of 0.54 [3]. CLC refers to the relatedness among a set of vertices, providing a global measure of relationships among all vertices. A good correlation between BWC and CLC is valuable when it comes to a connection between global and local view.

5. CONCLUSIONS

In this paper, we analyzed the six commonly studied centrality measures on random networks, small-world networks, and multiple real-world networks. A clear correlation of BWC to DEG is shown on most tested networks. It is safe to conclude that there is a strong correlation between BWC and DEG. In addition, FRC, CLC and EVC can also serve as alternatives to BWC in random network. For small-world networks, DEG, FRC, CLC and CCC could be preferably used as alternative to BWC at probability of rewiring lower than 0.2. Unlike the random and small-world networks, BWC is relatively less correlated to CLC and FRC on five tested real-world networks out of nine. DEG still is one of the best alternatives to BWC on real-world networks. In conclusion, we have found the computationally-cheap DEG as a good candidate to replace computationally-expensive BWC on most occasions.

REFERENCES

[1] L. C. Freeman, "A Set of Measures of Centrality based on Betweenness," Sociometry, vol. 40, no. 1, pp. 35-41, March 1977.

[2] J. M. Anthonisse, "The Rush in a Directed Graph," Stichting Mathematisch Centrum. Mathemaatische Besliskunde (BN 9/71), pp. 1-10, 1971.

[3] L. Leydesdorff, "Betweenness Centrality as an Indicator of the Interdisciplinarity of Scientific Journals," Journal of the American Society for Information Science and Technology, vol. 58, no. 9, pp. 1303-1319, 2007.

[4] M. Barthelemy, "Betweenness Centrality in Large Complex Networks," European Physical Journal B, vol. 38, pp. 163-168, 2004.

[5] U. Brandes, "A Faster Algorithm for Betweenness Centrality," The Journal of Mathematical Sociology, vol. 25, no. 2, pp. 163-177, 2001.

[6] M. E. J. Newman, "Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality," Physical Review E, vol. 64, 016132, June 2001.

[7] D. Bader, S. Kintali, K. Madduri and M. Mihail, "Approximating Betweenness Centrality," Algorithms and Models for the Web-Graph, Lecture Notes in Computer Science, vol. 4863, pp. 124-137, 2007.

[8] M. E. J. Newman, "A Measure of Betweenness Centrality based on Random Walks," Social Networks, vol. 27, pp. 39-54, 2005.

[9] S. P. Borgatti and M. G. Everett, "A Graph-Theoretic Perspective on Centrality," Social Networks, vol. 28, no. 4, pp. 466-484, October 2006.

[10] N. Meghanathan, "A computationally lightweight and localized centrality metric in lieu of betweenness centrality for complex network analysis," Vietnam Journal of Computer Science, pp. 1-16, June 2016.

[11] M. E. J. Newman and M. Girvan, "Finding and Evaluating Community Structure in Networks," Physical Review E, vol. 69, 026113, February 2004.

[12] M. G. Everett and S. P. Borgatti, "The Centrality of Groups and Classes," The Journal of Mathematical Sociology, vol. 23, no. 9, pp. 181-201, 1999.

[13] K-I. Goh, E. Oh, B. Kahng and D. Kim, "Betweenness Centrality Correlation in Social Networks," Physical Review E, vol. 67, 017101, January 2003.

[14] K-I. Goh, B. Kahng and D. Kim, "Universal Behavior of Load Distribution in Scale-Free Networks," Physical Review Letters, vol. 87, 278701, December 2001.

[15] N. Meghanathan, "Distribution of Maximal Clique Size under the Watts-Strogatz Model of Evolution of Complex Networks," International Journal in Foundations of Computer Science and Technology, vol. 5, no. 3, pp. 1-12, May 2015.

[16] F. Pozzi, T. D. Matteo and T. Aste, "Spread of Risk across Financial Markets: Better to Invest in te Peripheries," Scientific Reports, vol. 3, no. 1665, 2013.