CIT. Journal of Computing and Information Technology, Vol. 25, No. 2, June 2017, 103–132
doi: 10.20532/cit.2017.1003492

Evaluation of Correlation Measures for Computationally-Light vs. Computationally-Heavy Centrality Metrics on Real-World Graphs

Natarajan Meghanathan

Department of Computer Science, College of Science, Engineering and Technology, Jackson State University, Jackson, USA

We identify three different levels of correlation (pair-wise relative ordering, network-wide ranking and prediction through linearity) that could be assessed between a computationally-light centrality metric and a computationally-heavy centrality metric for real-world networks. The Kendall's concordance-based correlation measure could be used to quantitatively assess how well we could consider the relative ordering of two vertices vi and vj with respect to a computationally-light centrality metric as the relative ordering of the same two vertices with respect to a computationally-heavy centrality metric. We hypothesize that the pair-wise relative ordering (concordance)-based assessment of the correlation between centrality metrics is the strictest of all three levels of correlation and claim that the Kendall's concordance-based correlation coefficient will be lower than the correlation coefficient observed with the more relaxed levels of correlation measures (prediction through linearity-based Pearson's product-moment correlation coefficient and the network wide ranking-based Spearman's correlation coefficient). We validate our hypothesis by evaluating the three correlation coefficients between two sets of centrality metrics: the computationally-light degree and local clustering coefficient complement-based degree centrality metrics and the computationally-heavy eigenvector centrality, betweenness centrality and closeness centrality metrics for a diverse collection of 50 real-world networks.

ACM CCS (2012) Classification: Computing methodologies → Modeling and Simulation → Simulation Theory → Network Science

Networks → Network properties → Network structure → Topology analysis and generation

Keywords: relative ordering, ranking, centrality, correlation

1. Introduction

Network Science deals with analyzing complex networks (e.g., biological networks, social networks, citation networks, web, etc) from a graph theoretic perspective [1]. We model a complex network as an abstract graph of vertices (nodes) and edges (links). Centrality of a vertex is a quantitative measure of the topological significance of the vertex in a graph [1]. There exists a slew of centrality metrics for complex network analysis. Among these, the commonly studied metrics are the degree-based degree centrality (DegC) [1] and eigenvector centrality (EVC) [2] metrics as well as the shortest path-based betweenness centrality (BWC) [3] and closeness centrality (CLC) metrics [4]. The degree centrality of a vertex is a measure of the number of neighbors of the vertex. The eigenvector centrality [2] of a vertex is a measure of the degree of the vertex as well as the degree of its neighbors. A vertex has a higher EVC if it has a high-degree and its neighbors also have a high-degree. The betweenness centrality [3] of a vertex is a measure of the number of shortest paths (between any two vertices in the network) that go through the vertex. The closeness centrality [4] of a vertex is a measure of the hop count of the shortest paths (or the weight of the shortest paths in a weighted graph) from the vertex to the rest of the vertices in a graph. For graphs that are not connected, the centrality metrics are typically computed for the largest connected component of the graph.

Among the above four centrality metrics (see Section 5 for a comparison of the computation time), the degree centrality metric is the only computationally-light metric (i.e., can be computed quickly) and the other three metrics are computationally-heavy (i.e., will take more computation time). For a graph of V vertices and E edges, it takes $\Theta(V^3)$ time to compute the EVC of the vertices [2], and $\Theta(V^2 + VE)$ time to compute each of the BWC [3] and CLC metrics [4]. Recently, some research articles (e.g., [5] – [6]) have evaluated the correlation between these four commonly used centrality metrics for real-world network graphs to see if one or more of the computationally-heavy centrality metrics (EVC, BWC, CLC) exhibit a strong correlation with the degree centrality metric (on the basis of the Pearson's correlation coefficient [7]) so that one could then employ linear regression to predict the strongly correlated computationally-heavy metric(s) using the degree centrality metric. On similar lines, the Pearson's correlation coefficient between each of the above four centrality metrics and the maximal clique size per node (another node-level computationally-heavy metric, the computation of which is a NP-hard problem) was evaluated in [8].

In a recent work [9], a new metric called the localized clustering coefficient complement-based degree centrality (LCC'DC) has been proposed as a computationally-light alternative to the computationally-heavy BWC metric. LCC'DC is computed as the product of 1 – LCC and DegC; where LCC (local clustering coefficient) of a vertex [1] is a measure of the probability that any two neighbors of the vertex are connected and is computed as the ratio of the actual number of edges between the neighbors of a vertex to that of the maximum possible number of edges between the neighbors of the vertex. For several real-world networks analyzed in [9], the Pearson's correlation coefficient values observed for LCC'DC-BWC are larger than the correlation coefficient values observed for DegC-BWC. In another recent work [10], it was observed that compared to DegC, LCC'DC could be used to more accurately predict BWC values using linear regression (with standard error of residual values smaller than those incurred for regression using DegC). For all correlation analysis in this paper, we consider DegC and LCC'DC as the two computationally-light centrality metrics and EVC, BWC and CLC as the three computationally-heavy centrality metrics.

In this paper, we identify three different levels of correlation that could be evaluated between any two centrality metrics of the vertices (more specifically, between a computationally-light metric and a computationally-heavy metric) in complex network graphs:

(i) A pair-wise relative ordering-based correlation that would be a quantitative measure of how well the relative ordering of a pair of vertices based on a computationally-light metric could be considered as the relative ordering of the same pair of vertices with respect to a computationally-heavy metric. For example: if LCC'DC(v_i) < LCC'DC(v_j), how sure are we to say BWC(v_i) < BWC(v_j) for some two vertices v_i and v_j?

(ii) A network wide ranking-based correlation that would be a quantitative measure of the extent we could use the ranking of the vertices based on a computationally-light metric as the ranking of the vertices based on a computationally-heavy metric.

(iii) A linear regression-based correlation that would be a quantitative measure of the extent we could use the values of a computationally-light metric to predict the values for a computationally-heavy metric.

The Pearson's product-moment correlation coefficient is not the only correlation measure used in statistical analysis. There are at least two other well-known correlation measures such as the Spearman's Rank-based correlation measure [5] and the Kendall's concordance-based correlation measure [5] that are widely used in statistical analysis, but not that commonly used in complex network analysis. We opine that the Kendall's correlation coefficient (rather than the Pearson's correlation measure) could be more apt to do pair-wise relative ordering of the vertices with respect to a computationally-heavy metric based on the values incurred for a computationally-light metric. Likewise, the Spearman's rank-based correlation coefficient could be an apt measure to decide whether a computationally-light metric could be used to rank the vertices in a graph in lieu of a computationally-heavy metric. We claim that real-world network graphs are more likely to incur different values for the correlation coefficient with respect to each of the above three correlation measures and the Pearson's correlation coefficient alone cannot be used to infer the nature of correlation between any two centrality metrics with respect to each of the three levels of correlation that are of interest in this paper. For example: for the US Politics Books Network [11], we observed the following values for the Kendall's, Pearson's and Spearman's correlation coefficients with respect to LCC'DC-BWC: 0.69, 0.78 and 0.86.

Our hypothesis in this paper is that the pairwise relative ordering-based correlation is the strictest of the three levels of correlation and the Kendall's correlation coefficient is more likely to be the lowest of the three correlation coefficients evaluated for real-world network graphs. This is because the correlation measure is quantified as the ratio of the difference between the number of concordant pairs and the number of discordant pairs to that of the total number of pairs of vertices. A pair of vertices v_i and vj are said to be concordant with respect to centrality metrics X and Y, if $\{X(v_i) < X(v_j)$ and $Y(v_i) < Y(v_j)\}$ or $\{X(v_i) > X(v_j)$ and $Y(v_i) > Y(v_j)\}$ or $\{X(v_i) = X(v_j)$ and $Y(v_i) = Y(v_j)\}$; and discordant if $\{X(v_i) < X(v_j)$ and $Y(v_i) > Y(v_j)\}$ or $\{X(v_i) > X(v_j)$ and $Y(v_i) < Y(v_j)\}$. The Kendall's concordance-based correlation is evaluated at the level of vertex-vertex pairs and hence for two centrality metrics to be strongly correlated according to this measure, the number of concordant pairs of vertices should be significantly larger than the number of discordant pairs of vertices. The presence of even few discordant pairs of vertices could significantly reduce the value for the Kendall's correlation coefficient. For two different centrality metrics: if the number of concordant pairs of vertices is significantly larger than the number of discordant pairs of vertices, the network-wide ranking of the vertices with respect to the two centrality metrics is expected to be more or less the same. Likewise, the larger the number of concordant pairs of vertices with respect to two centrality metrics X and Y, the larger the chances of a dependence of the values for the centrality metric Y on the values for centrality metric X and vice-versa. Unless the centrality value of a vertex with respect to metric Y increases (or decreases) with an increase (or decrease) in the centrality value of the vertex with respect to metric X, it would be difficult to find a significant number of concordant pairs of vertices with respect to the two centrality metrics X and Y. Hence, we claim that the correlation coefficient between two centrality metrics for a real-world network graph could be bounded below by the Kendall's concordance-based correlation coefficient. In other words, if we could evaluate the Kendall's concordance-based correlation coefficient between two centrality metrics for a real-world network graph, the correlation coefficients expected between the same two centrality metrics with respect to the other two correlation measures (i.e., the Spearman's and Pearson's measures) are more likely to be at least the value obtained for the Kendall's concordance-based correlation coefficient.

We determine the correlation coefficient for DegC and LCC'DC with each of the three computationally heavy centrality metrics (EVC, BWC and CLC) with respect to the three different measures of correlation for a total of 50 real-world networks. This generates a huge dataset of correlation coefficient values (50 networks · 2 computationally-light metrics: DegC and LCC'DC · 3 computationally-heavy metrics: EVC, BWC and CLC = 300 combinations) for each of the three correlation measures. We determine the fraction of the combinations for which each of the three correlation coefficient measures incurs the lowest and largest values. We observe the Kendall's concordance-based correlation coefficient to be the lowest for 75% of the combinations: this strongly indicates the validity of the hypothesis.

Throughout the paper, the terms "network" and "graph", "node" and "vertex", "link" and "edge" are used interchangeably; they mean the same. All the real-world network graphs and the example graphs analyzed in this paper are modeled as undirected graphs. The adjacency matrix of an undirected graph of V vertices is a $V \times V$ binary matrix wherein there is an entry of 1 for cells (v_i, v_j) and (v_j, v_i) if and only if there is an edge between the two vertices v_i and v_j; otherwise, the entry is a 0. The rest of the paper is organized as follows: In Section 2, we review the five centrality metrics DegC, LCC'DC, EVC, BWC and CLC, and illustrate their computation with an example graph. In Section 3, we review the three correlation measures (Kendall's, Spearman's and Pearson's) and illustrate their computation for a computationally-light metric vs. a computationally-heavy metric for the example graph in Section 2. In Section 4, we present the 50 real-world networks analyzed in this paper and tabulate the values for some of the fundamental metrics. We also tabulate the computation time for the five centrality metrics (on the 50 real-world networks) justifying their classification as computationally-light or computationally-heavy. In Section 5, we present the results of the correlation analysis conducted on the 50 real-world networks on the basis of computationally-light vs. computationally-heavy centrality metrics with respect to the three correlation measures. In Section 6, related work is presented and unique contributions of the work are highlighted. Section 7 concludes the paper.

2. Review of Centrality Metrics

Centrality metrics quantify the importance of a vertex with respect to its position in a graph. In this paper, we consider centrality metrics on the basis of whether they are computationally-light or computationally-heavy. We identify the degree centrality (DegC) [1] and the recently proposed localized clustering coefficient complement-based degree centrality (LCC'DC) [9] as the two computationally-light centrality metrics (as they could be computed quickly with time; see Section 4) and identify the other three well-known centrality metrics: eigenvector centrality (EVC) [2], betweenness centrality (BWC) [3] and closeness centrality (CLC) [4] as the computationally-heavy metrics. In this section, we review each of these five metrics and illustrate their computation with a running example graph.

2.1. Degree Centrality

The degree centrality (DegC) of a vertex is the number of neighbors incident on the vertex. Figure 1 illustrates the degree centrality of the vertices (listed above the vertices) in the example graph used in Sections 2 and 3. A key weakness of the degree centrality metric is that the metric can take only integer values and ties among vertices (with same degree) is quite common and unavoidable in network graphs of any size (in the graph of Figure 1, we observe five of the nine vertices to have a degree of 3). Due to this inherent weakness, we opine that degree centrality might not be an apt metric for network-wide ranking of the vertices or pair-wise relative ordering of the vertices in lieu of the computationally-heavy metrics, even though DegC has been observed [5] – [6] to be strongly correlated with the computationally-heavy centrality metrics (EVC, BWC and CLC) with respect to the Pearson's correlation measure for linear dependence.

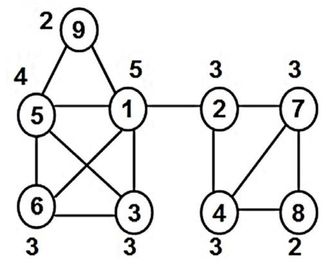


Figure 1. Degree centrality of the vertices in an example graph.

2.2. Eigenvector Centrality

The eigenvector centrality (EVC) of a vertex is a measure of the degree of the vertex as well as the degree of its neighbors [2]. The EVC of the vertices is a column vector computed using the power-iteration algorithm [13]. The algorithm takes the adjacency matrix of the graph (say, $A[v_i, v_j]$ for $1 \leq v_i, v_j \leq V$, where V is the number of vertices) as input and processes it through a sequence of iterations. The EVC column vector is initialized to a unit vector (all the entries are 1 s). In the first iteration, we multiply the adjacency matrix A with the EVC column vector of all 1 s and divide the entries in the product vector P (also a column vector) by the normalized value of its entries. The normalized value for a vector is the square root of the sum of the squares of the entries in the vector. For the subsequent iterations, we set the EVC column vector to be the product vector obtained (after dividing the individual entries with the normalized value) at the end of the previous iteration. We continue the iterations by multiplying the adjacency matrix with the EVC column vector obtained at the end of the previous iteration. The algorithm stops when the entries in the EVC column vector are close enough (i.e., do not change further to a certain level of precision) and the vector is then called the principal eigenvector.

There is an entry for each vertex in the principal eigenvector and the values of these entries correspond to the eigenvector centrality of the vertices. The normalized value of the entries in the final product vector that is transformed to the principal eigenvector is called the principal eigenvalue (a.k.a. the spectral radius) of the adjacency matrix of the network graph. The power-iteration method is of time-complexity $\Theta(V^3)$ as we do $\Theta(V^2)$ multiplications in each iteration (to compute the product vector) and there could be at most V iterations before the entries in the product vector converge and the product vector becomes the principal eigenvector. However, if the real-world network graph is a sparse graph, several state-of-the-art approaches (like [63]) for sparse matrix-vector multiplication (see [64] for a recent survey of the available approaches) could be employed to reduce the time-complexity that would only depend on the number of non-zero entries in the adjacency matrix.

Figure 2 presents an example to illustrate the computation of the principal eigenvector (i.e., the EVC of the vertices) for the example graph. The example aptly illustrates the impact of the DegC and EVC of the neighbors of a vertex on the EVC of the vertex. We notice that though vertices 8 and 9 have the same degree of 2, vertex 9 has a relatively larger EVC (0.290) compared to vertex 8 (0.069): this is because, vertex 9 is attached to two neighboring vertices (vertices 1 and 5) that have a larger DegC as well as a larger EVC; whereas, vertex 8 is attached to two neighboring vertices (vertices 4 and 7) that have a relatively lower DegC and lower EVC values.

2.3. Betweenness Centrality

The betweenness centrality (BWC) of a vertex is a measure of the number of shortest paths between any two vertices that go through the vertex [3]. The BWC of a vertex v_i is quantitatively computed as follows:

$$BWC(v_i) = \sum_{\substack{v_j \neq v_i \\ v_k \neq v_i}} \frac{\# sp_{v_i}(v_j, v_k)}{\# sp(v_j, v_k)},$$

where $\# sp\,(v_j, v_k)$ is the total number of shortest paths between any two vertices v_j and v_k (other than v_i) and $\# sp\,(v_j, v_k)$ is the number of such shortest paths between vertices v_j and v_k that go through vertex v_i.

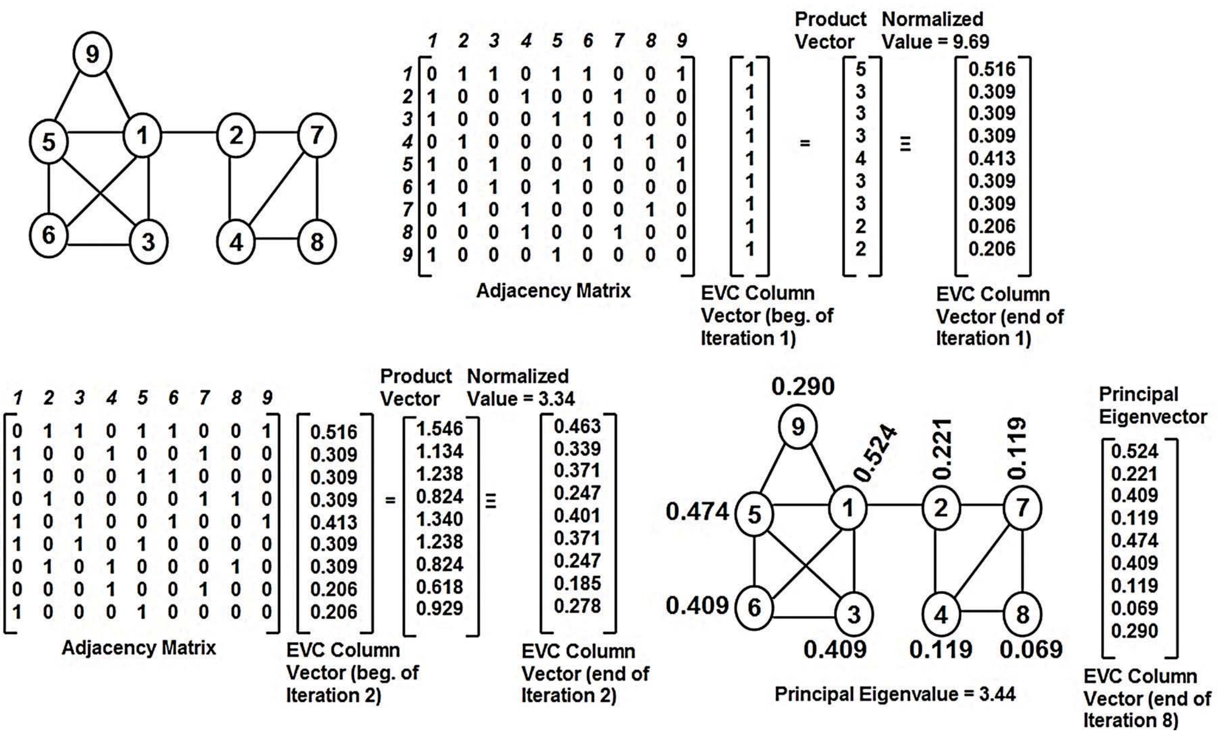


Figure 2. Eigenvector centrality of the vertices in an example graph.

BWC is a computationally-heavy metric and the best algorithm known so far is the classical Brandes's algorithm [14] of time-complexity Θ $(V^2 + VE)$ for undirected graphs. We now briefly describe a breadth first search (BFS)-based implementation [15] of the Brandes's algorithm. We compute a BFS tree rooted at each of the vertices in the graph; we keep track of the level number of every vertex (say, v_i in general) in each of these BFS trees. The level number of a vertex vi in a BFS tree rooted at vertex v_j corresponds to the number of hops on the shortest path from vertex v_j to v_i. One or more vertices could exist at a particular level in the BFS trees; a vertex v_x is considered to be a predecessor for a vertex v_y in a BFS tree if there exists an edge between v_x and v_y and v_x is at a level one less than the level of v_y (i.e., v_x is relatively closer to the root of the BFS tree). The root of a BFS tree is considered to be at level 0 for the particular tree. The number of shortest paths from the root of a BFS tree to itself is 1. The number of shortest paths for a vertex v_i from the root v_j of a BFS tree is the sum of the number of shortest paths from the root v_j to each of the predecessors of v_i in the BFS tree rooted at v_j. By using the level numbers and the set of predecessors of a vertex in a BFS tree rooted at a vertex v_j, we could calculate the number of shortest paths from the root v_j to every other vertex in the graph. To calculate the number of shortest paths from two vertices v_j to v_k that go through vertex v_i, we would simply take the maximum of the number of shortest paths from v_j to v_i (on the BFS tree rooted at v_j) and the number of shortest paths from v_k to v_i (on the BFS tree rooted at v_k). We can calculate the ratio

$$BWC(v_i; v_j, v_k) = \frac{\#sp_{v_i}(v_j, v_k)}{\#sp(v_j, v_k)},$$

for every vertex v_i with respect to the pair of vertices v_j and v_k ($v_j \neq v_i$ and $v_k \neq v_i$) and add these ratios to calculate the BWC of a vertex v_i. Figure 3 illustrates the computation of the BWC of the vertices in the example graph of

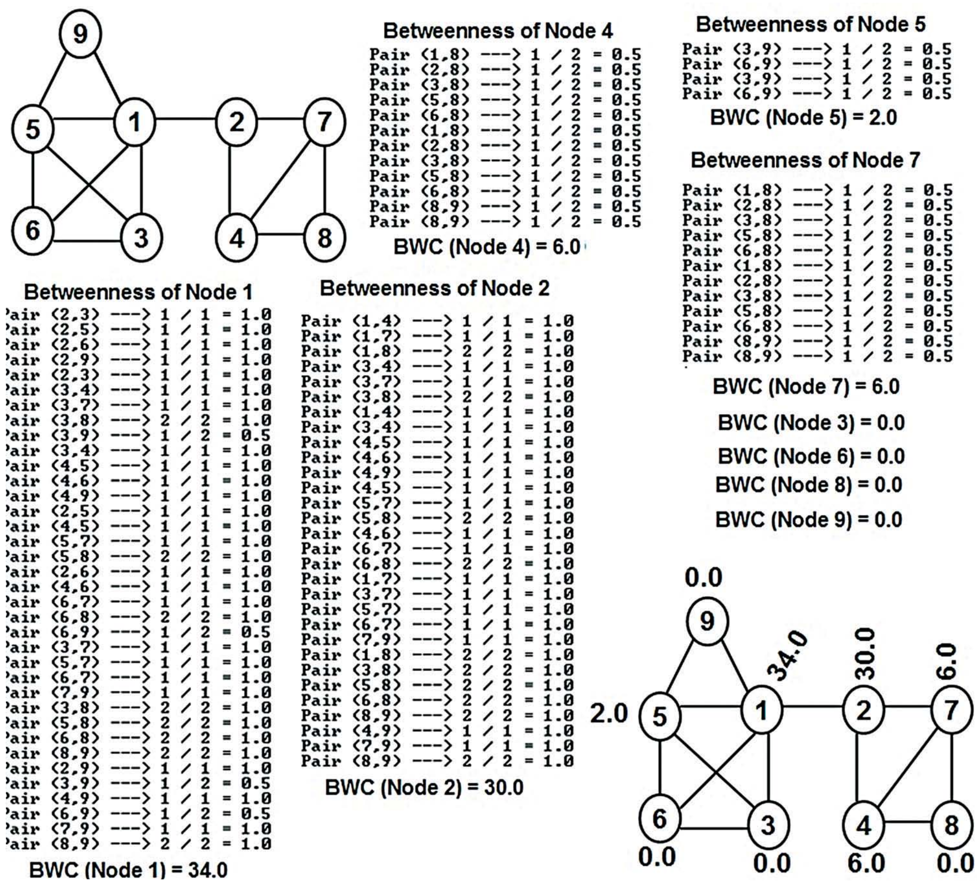
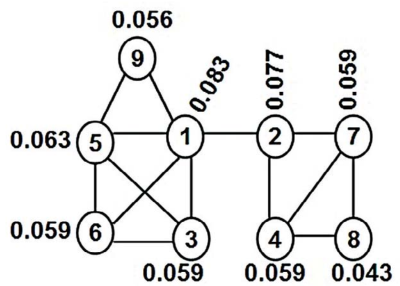


Figure 3. Betweenness centrality of the vertices in an example graph.

Figures 1 – 2. To avoid cluttering in the figure, we only show the non-zero BWC fractions of a vertex with respect to the pairs of vertices.

2.4. Closeness Centrality

The closeness centrality (CLC) of a vertex [4] is a measure of the closeness of the vertex to the rest of the vertices in a graph. The CLC of a vertex is computed as the inverse of the sum of the hop counts of the shortest paths from the vertex to the rest of the vertices in the graph. To determine the CLC of a vertex, we could use the $\Theta(V+E)$-BFS algorithm to determine a shortest path tree rooted at the vertex and find the sum of the level numbers of the vertices on this shortest path tree. We want to maintain the convention that the larger the centrality value for a vertex, the more important the vertex is. Hence, we find the inverse of the final sum of the level numbers of the vertices on the BFS-tree of a vertex and use it as the CLC of the vertex (rather than using just the sum of the level numbers as the CLC). Since we need to run the BFS algorithm once for each vertex, the overall time complexity to determine the CLC of the vertices is $\Theta(V(V+E)) = \Theta(V^2 + VE)$. Figure 4 illustrates the distance matrix (hop counts of the shortest paths between any two vertices) for the example graph of Figures 1 – 3 and also displays the CLC of the vertices. Vertex 1 is the closest vertex to the rest of the vertices (sum of the distances is 12, the minimum) and hence has the largest CLC value of 1/12 = 0.083.



	1	2	3	4	5	6	7	8	9	Sum of the Distances	CLC (Inverse of the Sum of the Distances)
1	0	1	1	2	1	1	2	3	1	12	0.083
2	1	0	2	1	2	2	1	2	2	13	0.077
3	1	2	0	3	1	1	3	4	2	17	0.059
4	2	1	3	0	3	3	1	1	3	17	0.059
5	1	2	1	3	0	1	3	4	1	16	0.063
6	1	2	1	3	1	0	3	4	2	17	0.059
7	2	1	3	1	3	3	0	1	3	17	0.059
8	3	2	4	1	4	4	1	0	4	23	0.043
9	1	2	2	3	1	2	0	3	4	18	0.056

Distance Matrix

Figure 4. Closeness centrality of the vertices in an example graph.

2.5. Localized Clustering Coefficient-Complement based Degree Centrality

The localized clustering coefficient (LCC) of a vertex is a measure of the probability for any two neighbors of the vertex to be connected [1]. The LCC of a vertex is computed as the ratio of the actual number of links between the neighbors of the vertex to that of the maximum possible number of links between the neighbors of the vertex [1]. The LCC of a vertex ranges from 0.0 to 1.0. If any two neighbors of a vertex are directly connected to each other, then the LCC of the vertex is 1.0. On the other hand, if no two neighbors of a vertex have a link between them, then the LCC of the vertex is 0.0. Note that the LCC of a vertex vi with just one neighbor is 1.0 as the neighbor is connected to itself and need not go through the vertex v_i to reach itself.

If two neighbors v_j and v_k of a vertex v_i are not directly connected to each other, then it is more likely that the two vertices would use vertex v_i for shortest path communication. The larger the fraction of the pairs of neighbors of a vertex that are not directly connected to each other (i.e., lower the LCC of a vertex), the larger the chances for several of the neighbors of the vertex to go through the vertex for shortest path communication. This observation leads to the proposal of a new centrality metric [9] called the local clustering coefficient complement-based degree centrality (LCC'DC). The local clustering coefficient complement (LCC' = 1 – LCC) essentially captures the probability that any two neighbors of a vertex would go through the vertex for shortest path communication. The LCC'DC of a vertex is simply the product of LCC' and DegC, the degree centrality of the vertex.

The hypothesis behind the proposal for LCC'DC is that the larger the number of neighbors for a vertex and the larger the fraction of pairs of these neighbors going through the vertex for shortest path communication, the larger are the chances of the vertex having a higher BWC. It has been observed in [9] that LCC'DC is strongly correlated to BWC for a suite of 18 real-world networks of different domains (a subset of the networks analyzed in this paper) with wide-ranging variations in degree distribution. Note that to compute the BWC of even a single vertex, we would need to determine the shortest path trees rooted at every vertex. On the other hand, LCC'DC is a computationally-light metric that could be computed simply on the basis of the two-hop neighborhood of a vertex. Figure 5 illustrates the computation of the LCC'DC of the vertices in the example graph of Figures 1 – 4. A comparison of the BWC and the LCC'DC values incurred for the example graph in Figures 3 and 5 indicates a strong correlation between the two metrics. Vertices 1 and 2 are the top two vertices to have the largest BWC values of 34 and 30 respectively; vertices 1 and 2 are also the top two vertices to have the largest LCC'DC values of 3.0 and 2.0 respectively. Likewise, the BWC of vertices 3, 6, 8 and 9 are 0.0 each and the LCC'DC values of these vertices are also 0.0 each.

3. Levels of Correlation and the Correlation Measures

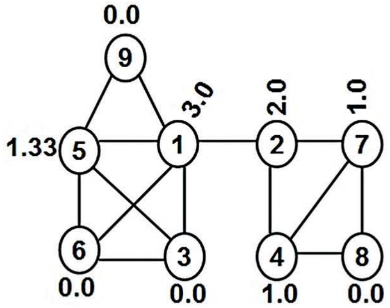
We identify three different levels of correlation that could be explored between a computationally-light centrality metric and computationally-heavy centrality metric. For discussion purposes, let X be a computationally-light centrality metric and Y be a computationally-heavy centrality metric. The three levels of correlation that are of interest in this paper are as follows:

(i) Pair-wise Relative Ordering of the Vertices: For any two vertices v_i and v_j, we are interested to quantify how well we can use the relative ordering of the two vertices with respect to the computationally-light metric X (i.e., whether $X(v_i) < X(v_j)$ or $X(v_i) > X(v_j)$ or $X(v_i) = X(v_j)$) as the relative ordering of the same two vertices with respect to the computationally-heavy metric Y.

(ii) Network-wide Ranking of the Vertices: We are interested to quantify how well we can use the network-wide ranking of the vertices with respect to a computationally-light metric X as the network-wide ranking of the vertices with respect to a computationally-heavy metric Y.

(iii) Predicting the Actual Centrality Values: We are interested to quantify how well we can predict the actual centrality values for the vertices with respect to a computationally-heavy metric Y based on the actual centrality values for the vertices with respect to a computationally-light metric X.

The Pearson's product-moment based correlation measure (r) has been the commonly used measure in the literature (e.g., [5] – [6]) to assess the correlation between centrality metrics for complex networks. However, the Pearson's correlation measure can accurately capture only one of the three levels of correlation (i.e., predicting the actual centrality values) and not the other two levels of correlation. The Kendall's concordance-based correlation measure (τ) and the Spearman's rank-based correlation measure (ρ) are the correlation measures that can effectively capture the pair-wise relative ordering of the vertices and the network-wide ranking of the vertices respectively. This is quite evident from their formulation itself (as will be seen in this section). The values for all three correlation coefficient measures range from –1 to 1; the closer the value to 1 or –1, the stronger (positive or negative) the correlation between



Vertex	Degree DegC	Max. Edges among neighbs	Actual Edges among neighbs	LCC	LCC' 1-LCC	LCC'DC
1	5	5(5-1)/2 = 10	4	0.4	0.6	3.0
2	3	3(3-1)/2 = 3	1	0.33	0.67	2.0
3	3	3(3-1)/2 = 3	3	1.0	0.0	0.0
4	3	3(3-1)/2 = 3	2	0.67	0.33	1.0
5	4	4(4-1)/2 = 6	4	0.67	0.33	1.33
6	3	3(3-1)/2 = 3	3	1.0	0.0	0.0
7	3	3(3-1)/2 = 3	2	0.67	0.33	1.0
8	2	2(3-1)/2 = 1	1	1.0	0.0	0.0
9	2	2(2-1)/2 = 1	1	1.0	0.0	0.0

Figure 5. LCC'DC of the vertices in an example graph.

the two centrality metrics in consideration. If the value for the correlation coefficient is closer to 0, it implies that there is no tendency for a centrality metric to either decrease or increase as the other increases.

Our hypothesis in this paper is that the pair-wise relative ordering of the vertices is the most restrictive level of correlation one could impose to assess the correlation among vertices with respect to any node-level metric (like centrality metric) and hence the Kendall's concordance-based correlation coefficient has a higher chance of being the lowest of the correlation coefficient values (compared to Pearson's r and Spearman's ρ) for the three levels of correlation. On the other hand, we conjecture that the Spearman's rank-based correlation is the least restrictive of the three correlation measures as it does not require the two centrality metrics to have a linear dependence (as is required for the Pearson's correlation measure) and minor differences in the rank of a vertex with respect to the two centrality metrics in consideration do not significantly affect the value for the correlation.

3.1. Kendall's Concordance-Based Correlation

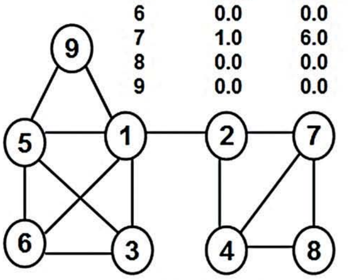
A pair of vertices v_i and v_j is said to be concordant with respect to centrality metrics X and Y if $\{X(v_i) < X(v_j)$ and $Y(v_i) < Y(v_j)\}$ or $\{X(v_i) > X(v_j)$ and $Y(v_i) > Y(v_j)\}$ or $\{X(v_i) = X(v_j)$ and $Y(v_i) = Y(v_j)\}$. A pair of vertices v_i and v_j is said to be discordant with respect to centrality metrics X and Y if $\{X(v_i) < X(v_j)$ and $Y(v_i) > Y(v_j)\}$ or $\{X(v_i) > X(v_j)$ and $Y(v_i) < Y(v_j)\}$. The Kendall's concordance [5]-based correlation coefficient (τ) is computed (see formulation 1) as the ratio of the difference between the number of concordant pairs ($\#\,conc.pairs$) and the number of discordant pairs ($\#\,disc.pairs$) to that of the total number of pairs of vertices (which is also the sum of the number of concordant pairs and discordant pairs).

$$\tau(X,Y) = \frac{\#\,conc.pairs(X,Y) - \#\,disc.pairs(X,Y)}{\#\,conc.pairs(X,Y) + \#\,disc.pairs(X,Y)} \tag{1}$$

Figure 6 illustrates an example on how to calculate the Kendall's concordance-based correlation between the LCC'DC and BWC metrics. We use the alphabets "C" (for concordance) and "D" (for discordance) to indicate whether a pair of vertices is concordant or discordant. In this example graph, there is a total of $9 \cdot (9-1)/2 = 36$ pairs of vertices that could be tested for concordance. Except the two pairs of vertices: pairs $4-5$ and $5-7$, all the other 34 pairs of vertices are observed to be concordant with respect to LCC'DC and BWC. Hence, the Kendall's concordance-based correlation coefficient $\tau(\text{LCC'DC}, \text{BWC}) = (34-2)/36 = 0.89$.

Note that due to the nature of the formulation (# concordant pairs – # discordant pairs) in the numerator, Kendall's concordance-based

Vertex	LCC'DC	BWC
1	3.0	34.0
2	2.0	30.0
3	0.0	0.0
4	1.0	6.0
5	1.33	2.0
6	0.0	0.0
7	1.0	6.0
8	0.0	0.0
9	0.0	0.0



τ (LCC'DC, BWC)
= (34 – 2)/36
= 0.89

Pair V_i, V_j	LCC'DC (V_i, V_j)	BWC (V_i, V_j)	C/D
1, 2	3.0, 2.0	34.0, 30.0	C
1, 3	3.0, 0.0	34.0, 0.0	C
1, 4	3.0, 1.0	34.0, 6.0	C
1, 5	3.0, 1.33	34.0, 2.0	C
1, 6	3.0, 0.0	34.0, 0.0	C
1, 7	3.0, 1.0	34.0, 6.0	C
1, 8	3.0, 0.0	34.0, 0.0	C
1, 9	3.0, 0.0	34.0, 0.0	C
2, 3	2.0, 0.0	30.0, 0.0	C
2, 4	2.0, 1.0	30.0, 6.0	C
2, 5	2.0, 1.33	30.0, 2.0	C
2, 6	2.0, 0.0	30.0, 0.0	C
2, 7	2.0, 1.0	30.0, 6.0	C
2, 8	2.0, 0.0	30.0, 0.0	C
2, 9	2.0, 0.0	30.0, 0.0	C
3, 4	0.0, 1.0	0.0, 6.0	C
3, 5	0.0, 1.33	0.0, 2.0	C
3, 6	0.0, 0.0	0.0, 0.0	C

Pair V_i, V_j	LCC'DC (V_i, V_j)	BWC (V_i, V_j)	C/D
3, 7	0.0, 1.0	0.0, 6.0	C
3, 8	0.0, 0.0	0.0, 0.0	C
3, 9	0.0, 0.0	0.0, 0.0	C
4, 5	1.0, 1.33	6.0, 2.0	D
4, 6	1.0, 0.0	6.0, 0.0	C
4, 7	1.0, 1.0	6.0, 6.0	C
4, 8	1.0, 0.0	6.0, 0.0	C
4, 9	1.0, 0.0	6.0, 0.0	C
5, 6	1.33, 0.0	2.0, 0.0	C
5, 7	1.33, 1.0	2.0, 6.0	D
5, 8	1.33, 0.0	2.0, 0.0	C
5, 9	1.33, 0.0	2.0, 0.0	C
6, 7	0.0, 1.0	0.0, 6.0	C
6, 8	0.0, 0.0	0.0, 0.0	C
6, 9	0.0, 0.0	0.0, 0.0	C
7, 8	1.0, 0.0	6.0, 0.0	C
7, 9	1.0, 0.0	6.0, 0.0	C
8, 9	0.0, 0.0	0.0, 0.0	C

Figure 6. Example to compute the Kendall's concordance-based correlation coefficient.

correlation coefficient has the tendency to reduce appreciably even in the presence of few discordant pairs. We more formally analyze the relationship between the number of concordant pairs and the number of discordant pairs on the Kendall's correlation coefficient as follows. Let $f_c(X, Y)$ be the fraction of the concordant pairs of vertices with respect to any two metrics X and Y; the formulation for Kendall's concordance-based correlation coefficient could be written as follows.

Fraction of concordant pairs of vertices,

$$f_c(X,Y) = \frac{\#conc.pairs(X,Y)}{\#conc.pairs(X,Y) + \#disc.pairs(X,Y)}$$

$$1 - f_c(X,Y) = \frac{\#conc.pairs(X,Y)}{\#conc.pairs(X,Y) + \#disc.pairs(X,Y)}$$

$$\tau(X,Y) = \frac{\#conc.pairs(X,Y)}{\#conc.pairs(X,Y) + \#disc.pairs(X,Y)} - \frac{\#disc.pairs(X,Y)}{\#conc.pairs(X,Y) + \#disc.pairs(X,Y)}$$

$$\tau(X,Y) = f_c\left(X,Y\right) - \left(1 - f_c\left(X,Y\right)\right)$$

$$\tau(X,Y) = 2 f_c\left(X,Y\right) - 1. \tag{2}$$

Figure 7 illustrates how $\tau(X, Y)$ decreases with decrease in $f_c(X, Y)$. We can notice that for a 0.01 decrease in $f_c(X, Y)$, $\tau(X, Y)$ decreases by 0.02.

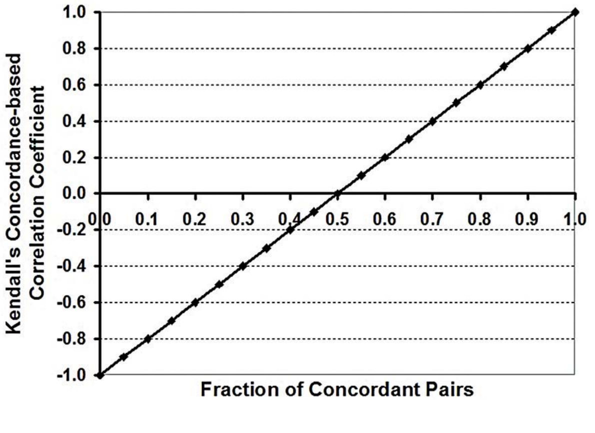


Figure 7. Relationship between Kendall's concordance-based correlation coefficient and the fraction of concordant pairs.

3.2. Spearman's Rank-Based Correlation

The rank of a vertex with respect to a centrality metric is a measure of where the vertex stands if the vertices in the network are to be ordered in the decreasing order of the values for the centrality metric (we assume decreasing order for all the centrality metrics). The earlier a vertex appears in the listing with respect to a particular centrality metric, the higher the rank for the vertex with respect to the metric. We use the Spearman's rank [5]-based correlation measure (ρ) to quantify the extent of similarity in the ranking of the vertices with respect to two centrality metrics. We calculate this correlation coefficient measure as follows with respect to any two centrality metrics (say X and Y). For each centrality metric: we first obtain a listing of the vertices in the decreasing order of the centrality values. If two or more vertices have the same centrality value, we break the tie in favor of the vertex with the smaller ID. The index at which a vertex appears in this list is the tentative ranking for the vertex. The final ranking for a vertex with respect to a centrality metric is the same as the tentative ranking for the vertex if it has no tie with any other vertex for the centrality metric. If two or more vertices have a tie with respect to a centrality metric, their final ranking with respect to the centrality metric is the average of the tentative rankings for the vertices with respect to the metric. Let d_i be the difference in the final ranking for the vertices with respect to the two centrality metrics X and Y, where $1 \le i \le n$ and n is the number of vertices in the graph. The Spearman's rank-based correlation coefficient is computed as follows:

$$\rho(X,Y) = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2 - 1\right)}. \tag{3}$$

Figure 8 illustrates the computation of the Spearman's rank-based correlation coefficient in the example graph of Figures 1 – 5. For BWC, we observe vertices 4 and 7 to have tie (BWC = 6 for both) and we break the tie on the basis of the vertex ID: vertex 4 with a lower ID gets a tentative rank of 3 and vertex 7 gets a tentative rank of 4; the final ranking for the two vertices is the average of their tentative rankings ($(3+4)/2 = 3.5$). A similar tie between the two vertices exists with respect to LCC'DC. We also observe the tie between vertices 3, 6, 8 and 9 with respect to both BWC and LCC'DC. We observe a non-zero difference in the ranking of the vertices for only three of the nine vertices and the magnitudes of these differences are not that high to significantly reduce the correlation coefficient value (0.95).

With respect to formulation (3): for larger values of n, the term in the denominator $n\,(n^2-1)$ dominates the summation term $\sum_{i=1}^{n} d_i^2$ in the numerator. Hence, even if the differences in the ranking of the vertices are larger, the Spearman's rank-based correlation coefficient is more likely to stay relatively high (compared to the Kendall's measure) for graphs with larger number of vertices.

3.3. Pearson's Product-Moment Correlation

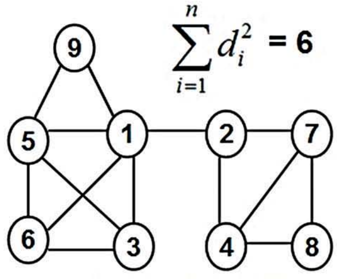
The Pearson's product-moment correlation (r) when applied for centrality metrics is a measure of the linear dependence between any two metrics in consideration [5]. It is referred to as the product-moment based correlation as we calculate the deviation of the data points from their mean value ("mean" is also referred to as 'first moment' in statistics) and use them in the formulation to calculate the correlation coefficient (see formulation (4)). If X and Y are the datasets for two centrality metrics: let X_i and Y_i indicate centrality values for the individual vertices v_i ($1 \le i \le n$, where n is the number of vertices) and and are the average of the centrality values; $r\,(X, Y)$ is calculated as follows.

$$r\left(X,Y\right)=\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)}{\sqrt{\sum_{i=n}^{n}\left(X_i-\overline{X}\right)^2}\sqrt{\sum_{i=n}^{n}\left(Y_i-\overline{Y}\right)^2}} \quad (4)$$

Figure 9 illustrates the computation of Pearson's product-moment correlation coefficient in the example graph of Figures 1 – 5. We observe the Pearson's correlation coefficient to be 0.91 and is in between the values of 0.89 and 0.95 observed respectively for the Kendall's and Spearman's correlation coefficients. As seen for several real-world networks analyzed in this paper, the Kendall's correlation coefficient measure is the lowest of the three correlation coefficient values.

4. Real-World Networks

In this section, we provide a brief description of the 50 real-world networks analyzed in this paper and tabulate the values for some of the fundamental metrics for complex network analysis observed for these networks as well as tabulate the computation time per node of the five centrality metrics (discussed in Section 2) for these networks. All the real-world networks are modeled as undirected graphs. Table 1 lists the number of nodes and edges in these graphs

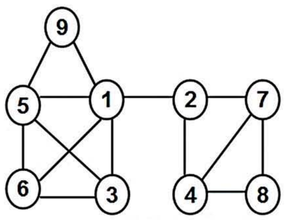


$$\sum_{i=1}^{n} d_i^2 = 6$$

$$\rho\,(\text{LCC'DC, BWC}) = 1 - \frac{(6*6)}{(9*(9^2-1))} = 0.95$$

Vertex, i	BWC	Tentative Rank: BWC	Final Rank: BWC_i	LCC'DC	Tentative Rank: LCC'DC	Final Rank: $LCC'DC_i$	Rank Difference d_i = BWC_i - $LCC'DC_i$	d_i^2
1	34	1	1	3	1	1	0	0
2	30	2	2	2	2	2	0	0
3	0	6	7.5	0	6	7.5	0	0
4	6	3	3.5	1	4	4.5	-1	1
5	2	5	5	1.33	3	3	2	4
6	0	7	7.5	0	7	7.5	0	0
7	6	4	3.5	1	5	4.5	-1	1
8	0	8	7.5	0	8	7.5	0	0
9	0	9	7.5	0	9	7.5	0	0

Figure 8. Example to compute the Spearman's rank-based correlation coefficient.



$$r(X,Y) = \frac{104.467}{\sqrt{1456} * \sqrt{9.059}} = 0.91$$

Vertex, i	BWC	BWC_i - Avg(BWC)	(BWC_i - Avg(BWC) $)^2$	LCC'DC	LCC'DC_i - Avg(LCC'DC)	(LCC'DC_i - Avg(LCC'DC) $)^2$	(BWC_i- Avg(BWC)* (LCC'DC_i - Avg(LCC'DC)
1	34	25.33	641.61	3	2.07	4.285	52.433
2	30	21.33	454.97	2	1.07	1.145	22.823
3	0	-8.67	75.17	0	-0.93	0.865	8.063
4	6	-2.67	7.13	1	0.07	0.005	-0.187
5	2	-6.67	44.49	1.33	0.4	0.160	-2.668
6	0	-8.67	75.17	0	-0.93	0.865	8.063
7	6	-2.67	7.13	1	0.07	0.005	-0.187
8	0	-8.67	75.17	0	-0.93	0.865	8.063
9	0	-8.67	75.17	0	-0.93	0.865	8.063
	Avg = 8.67		Sum = 1456	Avg = 0.93		Sum = 9.059	Sum = 104.467

Figure 9. Example to compute the Pearson's product-moment correlation coefficient.

as well as the values for fundamental metrics like average node degree (kavg), average path length (PL_{avg}), diameter (D), spectral radius ratio for node degree (λ_{sp}) [12], graph density (G_d), graph modularity (G_m, measured in a scale of 0...1) [16] and the number of components (# *comps*). The spectral radius ratio for node degree [12] is a measure of the variation in node degree and is calculated as the ratio of the principal eigenvalue [2] of the adjacency matrix of the graph to that of the average node degree. The spectral radius ratio for node degree is independent of the number of vertices and the actual degree values for the vertices in the graph. The spectral radius ratio for node degree is always greater than or equal to 1; the farther the ratio from the value of 1, the larger the variation in node degree. The spectral radius ratio for node degree for the real-world network graphs analyzed in this paper ranges from 1.01 to 5.34 (indicating that the real-world network graphs analyzed range from random networks with smaller variation in node degree to scale-free networks of larger variation in node degree).

The networks considered cover a broad range of categories (as listed below, along with the number of networks in each category): Acquaintance network (13), Friendship network (9), Co-appearance network (6), Employment network (4), Citation network (3), Collaboration network (3), Biological network (3), Political network (2), Game network (2), Literature network (2), Transportation network, Geographical network and Trade network (all 1 each). A brief description of each category of networks is as follows: An *acquaintance network* is a kind of social network in which the participant nodes slightly (not closely) know each other, as observed typically during an observation period. A *friendship network* is a kind of social network in which the participant nodes closely know each other and the relationship is not captured over an observation period. A *co-appearance network* is a network typically extracted from novels / books in such a way that two characters or words (modeled as nodes) are connected if they appear alongside each other. An *employment network* is a network in which the interaction / relationship between people is primarily due to their employment requirements and not due to any personal liking. A *citation network* is a network in which two papers (nodes) are connected if one paper cites the other paper as reference. A *collaboration network* is a network of researchers / authors who are listed as co-authors in at least one publication. A *biological network* is a network that models the interactions between genes, proteins, animals of a particular species, etc. A *political network* is a network of entities (typically politicians) involved in politics. A *game network* is a network of teams or players playing for different teams and their associations. A *literature network* is a network of papers / terminologies / authors (other than collaboration, citation or co-authorship) involved in a particular area of literature. A *transportation network* is a network of entities (like airports and their flight connections) involved in public transportation. A *geographical network* is a network of states and their shared borders in a country. A *trade network* is a network of countries / people involved in certain trade. More information about the individual real-world networks is given below:

Table 1. Fundamental metrics for real-world network graphs used in correlation analysis.

#	Net.	Network Type	λ_{sp}	#nodes	#edges	k_{avg}	D	G_d	G_m	#comps	PL_{avg}
1	ADJ	Co-appearance Net.	1.73	112	425	7.589	5	0.068	0.283	1	2.536
2	AKN	Co-appearance Net.	2.48	140	494	7.057	5	0.051	0.389	2	2.448
3	JBN	Employment Net.	1.45	198	2742	27.697	6	0.141	0.44	1	2.235
4	CEN	Biological Net.	1.68	297	2148	14.465	5	0.049	0.387	1	2.455
5	CLN	Citation Net.	2.03	118	613	10.39	4	0.089	0.297	1	2.374
6	CGD	Citation Net.	2.24	259	640	4.942	11	0.019	0.627	6	4.149
7	CFN	Co-appearance Net.	1.83	89	407	9.146	3	0.104	0.375	2	1.945
8	DON	Acquaintance Net.	1.40	62	159	5.129	8	0.084	0.521	1	3.357
9	DRN	Acquaintance Net.	2.76	212	284	2.679	18	0.013	0.734	9	7.03
10	DLN	Literature Net.	1.49	37	81	4.378	7	0.122	0.299	2	2.703
11	ERD	Collaboration Net.	3.00	433	1314	6.069	11	0.014	0.533	3	4.021
12	FMH	Friendship Net.	2.81	147	202	2.748	16	0.019	0.801	11	6.811
13	FHT	Friendship Net.	1.57	33	91	5.515	5	0.172	0.308	1	2.36
14	FTC	Employment Net.	1.21	48	170	7.083	5	0.151	0.462	1	2.402
15	FON	Game Net.	1.01	115	613	10.661	4	0.094	0.604	1	2.508
16	CDF	Acquaintance Net.	1.11	58	967	33.345	3	0.585	0.066	1	1.419
17	GD96	Citation Net.	2.38	180	228	2.533	8	0.014	0.65	1	4.417
18	MUN	Co-appearance Net.	2.54	167	301	3.605	9	0.022	0.807	20	3.879
19	GLN	Literature Net.	2.01	67	118	3.522	7	0.053	0.502	4	3.099
20	HTN	Acquaintance Net.	1.21	115	2164	37.635	3	0.33	0.095	2	1.662
21	HCN	Co-appearance Net.	1.66	76	302	7.947	4	0.106	0.546	4	2.142
22	ISP	Acquaintance Net.	1.69	309	1924	12.453	10	0.04	0.565	1	3.775
23	KCN	Acquaintance Net.	1.47	34	78	4.588	5	0.139	0.416	1	2.408
24	KFP	Acquaintance Net.	1.70	37	85	4.595	10	0.128	0.444	2	3.23
25	LMN	Co-appearance Net.	1.82	77	254	6.597	5	0.087	0.555	1	2.641
26	MDN	Biological Net.	1.04	62	1167	37.645	2	0.617	0.086	1	1.383
27	MTB	Acquaintance Net.	1.95	64	295	9.219	2	0.146	0.375	1	1.854
28	MCE	Employment Net.	1.12	77	1549	40.23	2	0.529	0.217	1	1.471
29	MSJ	Co-author Net.	3.48	475	625	2.632	17	0.006	0.945	104	6.49
30	AFB	Friendship Net.	2.29	171	940	10.994	7	0.065	0.688	4	3.069
31	MPN	Acquaintance Net.	1.23	35	117	6.686	4	0.197	0.357	1	2.106
32	MMN	Friendship Net.	1.59	30	61	4.067	5	0.14	0.424	1	2.644
33	NSC	Co-author Net.	5.51	1,589	2,743	3.45	17	0.002	0.959	269	5.823
34	PBN	Political Net.	1.42	105	441	8.4	7	0.081	0.525	1	3.079
35	PSN	Acquaintance Net.	1.22	238	5539	46.546	3	0.196	0.39	1	1.941
36	PFN	Friendship Net.	1.32	67	142	4.239	7	0.064	0.581	1	3.355
37	SJN	Acquaintance Net.	1.29	75	155	4.133	7	0.056	0.601	1	3.485
38	SDI	Employment Net.	1.94	230	359	3.122	14	0.014	0.696	5	5.607
39	SPR	Political Net.	1.57	92	477	10.37	5	0.114	0.25	1	2.32
40	SWC	Game Net.	1.45	35	118	6.743	5	0.198	0.231	1	2.123
41	SSM	Acquaintance Net.	1.22	24	38	3.167	6	0.138	0.562	1	2.993
42	TEN	Acquaintance Net.	1.06	22	39	3.545	5	0.169	0.444	1	2.494
43	TWF	Friendship Net.	1.49	47	77	3.277	8	0.071	0.741	4	2.652
44	UKF	Friendship Net.	1.35	83	578	13.928	4	0.17	0.45	2	2.097
45	APN	Transportation Net.	3.22	332	2126	12.807	6	0.039	0.358	1	2.738
46	USS	Geographical Net.	1.25	49	107	4.367	10	0.091	0.571	1	3.935
47	RHF	Friendship Net.	1.27	217	1839	16.949	4	0.078	0.426	1	2.395
48	WSB	Friendship Net.	1.22	43	336	15.628	3	0.372	0.255	1	1.671
49	WTN	Trade Net.	1.38	80	875	21.875	3	0.277	0.220	1	1.724
50	YPI	Biological Net.	3.20	1,870	2,203	2.387	19	0.001	0.841	149	6.810

Table 2. Computation time per node of the centrality metrics for the real-world networks.

#	Net.	#nodes	#edges	Computation Time per Node (milliseconds)				
				Computationally-light		Computationally-heavy		
				DegC	LCC'DC	CLC	EVC	BWC
1	ADJ	112	425	0.00043	0.00723	**0.09777**	**0.30250**	**2.40223**
2	AKN	140	494	0.00068	0.00965	**0.05050**	**0.44021**	**3.94329**
3	JBN	198	2742	0.00017	**0.04402**	**0.12066**	**0.22212**	**8.98010**
4	CEN	297	2148	0.00018	**0.01157**	**0.07825**	**0.47899**	**19.16182**
5	CLN	118	613	0.00023	0.00404	**0.05186**	**0.14542**	**1.45644**
6	CGD	259	640	0.00022	0.00083	**0.11286**	**0.48031**	**19.13170**
7	CFN	89	407	0.00017	0.00137	0.00674	**0.02854**	**0.46247**
8	DON	62	159	0.00018	0.00071	0.00419	**0.02097**	**0.31935**
9	DRN	212	284	0.00025	0.00058	**0.10759**	**0.27104**	**17.85425**
10	DLN	37	81	0.00027	0.00089	0.00216	**0.01919**	**0.12676**
11	ERD	433	1314	0.00019	0.00110	**0.20591**	**1.16956**	**48.16531**
12	FMH	147	202	0.00024	0.00050	**0.04871**	**0.12871**	**5.54497**
13	FHT	33	91	0.00024	0.00103	0.00182	**0.01485**	**0.13364**
14	FTC	48	170	0.00017	0.00079	0.00292	**0.01646**	**0.18542**
15	FON	115	613	0.00019	0.00121	**0.01330**	**0.08209**	**1.36739**
16	CDF	58	967	0.00028	0.00997	**0.01810**	**0.03414**	**0.67879**
17	GD96	180	228	0.00017	0.00052	**0.02817**	**0.09189**	**4.26378**
18	MUN	167	301	0.00018	0.00054	**0.02305**	**0.06587**	**1.50102**
19	GLN	67	118	0.00030	0.00046	0.00910	**0.03149**	**0.32149**
20	HTN	115	2164	0.00018	0.00724	**0.01165**	**0.05365**	**1.79522**
21	HCN	76	302	0.00026	0.00074	0.00855	**0.02579**	**0.32276**
22	ISP	309	1924	0.00017	0.00130	**0.10476**	**0.55414**	**21.06320**
23	KCN	34	78	0.00018	0.00047	0.00147	0.00529	**0.06882**
24	KFP	37	85	0.00030	0.00097	0.00324	**0.01216**	**0.16216**
25	LMN	77	254	0.00016	0.00083	0.00545	**0.01792**	**0.37195**
26	MDN	62	1167	0.00026	0.00560	0.00694	**0.03210**	**0.67774**
27	MTB	64	295	0.00017	0.00063	0.00500	**0.01609**	**0.32844**
28	MCE	77	1549	0.00017	0.00516	0.00558	**0.02377**	**0.74909**
29	MSJ	475	625	0.00020	0.00038	**0.18269**	**0.63120**	**28.86568**
30	AFB	171	940	0.00019	0.00137	**0.03135**	**0.38982**	**3.36468**
31	MPN	35	117	0.00017	0.00086	0.00171	0.00743	**0.10314**
32	MMN	30	61	0.00027	0.00047	0.00233	0.00700	**0.08767**
33	NSC	1,589	2,743	0.00016	0.00072	**2.52165**	**31.21962**	**457.18801**
34	PBN	105	441	0.00020	0.00092	**0.01848**	**0.07352**	**1.05924**
35	PSN	238	5539	0.00016	**0.01128**	**0.04836**	**0.31601**	**13.87235**
36	PFN	67	142	0.00016	0.00048	0.00567	**0.01925**	**0.33701**
37	SJN	75	155	0.00017	0.00055	0.00573	**0.02573**	**0.42813**
38	SDI	230	359	0.00018	0.00049	**0.05117**	**0.22422**	**10.97583**
39	SPR	92	477	0.00058	0.00133	**0.03793**	**0.14533**	**0.79196**
40	SWC	35	118	0.00017	0.00054	0.00143	0.00743	**0.08714**
41	SSM	24	38	0.00021	0.00033	0.00125	0.00417	**0.03292**
42	TEN	22	39	0.00018	0.00032	0.00091	0.00364	**0.03045**
43	TWF	47	77	0.00017	0.00032	0.00255	0.00979	**0.07106**
44	UKF	83	578	0.00016	0.00149	0.00675	**0.02675**	**0.62578**
45	APN	332	2126	0.00016	0.00323	**0.09518**	**0.49545**	**18.50593**
46	USS	49	107	0.00041	0.00045	0.00265	**0.01224**	**0.17469**
47	RHF	217	1839	0.00016	0.00212	**0.04083**	**0.24429**	**9.31433**
48	WSB	43	336	0.00019	0.00128	0.00209	0.00977	**0.17558**
49	WTN	80	875	0.00058	0.00283	**0.02513**	**0.10938**	**0.67938**
50	YPI	1,870	2,203	0.00018	0.00086	**3.45965**	**77.53588**	**834.37062**
Fraction of Networks for which Average Computation Time per Node ≥ 0.01 ms				0/50 = 0.0	3/50 = 0.06	26/50 = 0.52	42/50 = 0.84	50/50 = 1.00

1. Word Adjacency Network (ADJ; Newman, 2006b) [17]: This is a network of 112 words (adjectives and nouns, represented as vertices) in the novel *David Copperfield* by Charles Dickens; there exists an edge between two vertices if the corresponding words appeared adjacent to each other at least once in the novel.
2. Anna Karnenina Network (AKN; Knuth, 1993) [18]: This is a network of 140 characters (vertices) in the novel *Anna Karnenina*; there exists an edge between two vertices if the corresponding characters appeared together in at least one scene in the novel.
3. Jazz Band Network (JBN; Geiser & Danon, 2003) [19]: This is a network of 198 Jazz bands (vertices) that recorded between the years 1912 and 1940; there exists an edge between two bands if they shared at least one musician in any of their recordings during this period.
4. C. Elegans Neural Network (CEN; White *et al.*, 1986) [20]: This is a network of 297 neurons (vertices) in the neural network of the hermaphrodite *Caenorhabditis Elegans*; there is an edge between two vertices if the corresponding neurons interact with each other (in the form of chemical synapses, gap junctions and neuromuscular junctions).
5. Centrality Literature Network (CLN; Hummon *et al.*, 1990) [21]: This is a network of 118 papers (vertices) published on the topic of centrality in complex networks from 1948 to 1979. There is an edge between two vertices v_i and v_j if one of the corresponding papers has cited the other paper as a reference.
6. Citation Graph Drawing Network (CGD; Biedl & Franz, 2001) [22]: This is a network of 259 papers (vertices) that were published in the Proceedings of the Graph Drawing (GD) conferences from 1994 to 2000 and cited in the papers published in the GD'2001 conference. There is an edge between two vertices v_i and v_j if one of the corresponding papers has cited the other paper as a reference.
7. Copperfield Network (CFN; Knuth, 1993) [18]: This is a network of 89 characters in the novel *David Copperfield* by Charles Dickens; there exists an edge between two vertices if the corresponding characters appeared together in at least one scene in the novel.
8. Dolphin Network (DON; Lusseau *et al.*, 2003) [23]: This is a network of 62 dolphins (vertices) that lived in the Doubtful Sound fiord of New Zealand; there is an edge between two vertices if the corresponding dolphins were seen moving with each other during the observation period.
9. Drug Network (DRN; Lee, 2004) [24]: This is a network of 212 drug agents (vertices) of different ethnicities. There is a link between two vertices if the corresponding agents know each other.
10. Dutch Literature 1976 Network (DLN; Nooy, 1999) [25]: This is a network of 37 Dutch literary authors and critics (vertices) in 1976; there exists an edge between two vertices v_i and v_j if the person corresponding to one of them is a critic who made a judgment (through a review or interview) on the literary work of the author corresponding to the other vertex.
11. Erdos Collaboration Network (ERD; Batagelj & Mrvar, 2006) [26]: This is a network of 433 authors (nodes) who have either directly published an article with Paul Erdos or through a chain of collaborators leading to Paul Erdos. There is an edge between two nodes if the corresponding authors have co-authored at least one publication.
12. Faux Mesa High School Friendship Network (FMH; Resnick *et al.*, 1997) [27]: This is a network of 147 students (vertices) at a high school community in the rural western part of US; there exists an edge between two vertices if the corresponding students are friends of each other.
13. Friendship Ties in a Hi-Tech Firm (FHT; Krackhardt, 1999) [28]: This is a network of 33 employees (vertices) of a small hi-tech computer firm that sells, installs and maintains computer systems; there exists an edge between two vertices v_i and v_j if the employee corresponding to either one of them considers the employee corresponding to the other vertex as a personal friend.

14. Flying Teams Cadet Network (FTC; Moreno, 1960) [29]: This is a network of 48 cadet pilots (vertices) at an US Army Air Forces flying school in 1943 and the cadets were trained in a two-seated aircraft; there exists an edge between two vertices v_i and v_j if the pilot corresponding to either one of them has indicated the pilot corresponding to the other vertex as a preferred partner with whom s / he likes to fly during the training schedules.
15. US Football Network (FON; Girvan & Newman, 2002) [30]: This is a network of 115 football teams (nodes) of US universities that played in the Fall 2000 season; there is an edge between two nodes if the corresponding teams have played against each other in the league games.
16. College Dorm Fraternity Network (CDF; Bernard *et al.*, 1980) [31]: This is a network of 58 residents (vertices) in a fraternity college at a West Virginia college; there exists an edge between two vertices if the corresponding residents were seen in a conversation at least once during a five day observation period.
17. GD'96 Network (GD96; Batagelj & Mrvar, 2006) [26]: This is a network of 180 AT&T and other WWW websites (vertices) that were cited in the proceedings of the Graph Drawing (GD) conference in 1996; there exists an edge between two vertices if the website corresponding to one of them has a link to the website corresponding to the other vertex.
18. Marvel Universe Network (MUN; Gleiser, 2007) [32]: This is a collaborative network of 167 characters (vertices) in the comic books published by the Marvel Universe publishing company; there exists an edge between two vertices if the corresponding characters had appeared together in at least one publication.
19. Graph and Digraph Glossary Network (GLN; Batagelj & Mrvar, 2006) [26]: This is a network of 67 terms (vertices) that appeared in the glossary prepared by Bill Cherowitzo on Graphs and Digraphs; there appeares an edge between two vertices if the term corresponding to one of them is used to describe the meaning of the term corresponding to the other vertex.
20. Hypertext 2009 Network (HTN; Isella *et al.*, 2011) [33]: This is a network of the face-to-face contacts of 115 attendees (vertices) of the ACM Hypertext 2009 conference held in Turin, Italy from June 29 to July 1, 2009. There exists an edge between two vertices if the corresponding conference visitors had face-to-face contact that was active for at least 20 seconds.
21. Huckleberry Coappearance Network (HCN; Knuth, 1993) [18]: This is a network of 76 characters (vertices) that appeared in the novel Huckleberry Finn by Mark Twain; there is an edge between two vertices if the corresponding characters had a common appearance in at least one scene.
22. Infectious Socio-patterns Network (ISP; Isella *et al.*, 2011) [33]: This is a network of 309 visitors (vertices) who visited the Science Gallery in Dublin, Ireland during Spring 2009. There exists an edge between two vertices if the corresponding visitors had a continuous face-to-face contact for at least 20 seconds when they participated in the *Infectious Socio-patterns* event (an electronic simulation of the spreading of an epidemic through individuals in close proximity) as part of an art science exhibition.
23. Karate Club Network (KCN; Zachary, 1977) [34]: This is a network of 34 members (nodes) of a Karate Club at a US university in the 1970s; there is an edge between two nodes if the corresponding members were seen interacting with each other during the observation period.
24. Korea Family Planning Network (KFP; Rogers & Kincaid, 1980) [35]: This is a network of 37 women (vertices) at a Mothers' Club in Korea; there exists an edge between two vertices if the corresponding women were seen discussing family planning methods during an observation period.
25. Les Miserables Network (LMN; Knuth, 1993) [18]: This is a network of 77 characters (nodes) in the novel *Les Miserables*; there exists an edge between two nodes if the corresponding characters appeared together in at least one of the chapters in the novel.

26. Macaque Dominance Network (MDN; Takahata, 1991) [36]: This is a network of 62 adult female Japanese macaques (monkeys; vertices) in a colony, known as the "Arashiyama B Group", recorded during the non-mating season from April to early October, 1976. There exists an edge between two vertices if a macaque corresponding to one of them was recorded to have exhibited dominance over the macaque corresponding to the other vertex.
27. Madrid Train Bombing Network (MTB; Hayes, 2006) [37]: This is a network of 64 suspected individuals and their relatives (vertices) reconstructed by Rodriguez using press accounts in the two major Spanish daily newspapers (El Pais and El Mundo) regarding the bombing of commuter trains in Madrid on March 11, 2004. There exists an edge between two vertices if the corresponding individuals were observed to have a link in the form of friendship, ties to any terrorist organization, co-participation in training camps and/or wars, or co-participation in any previous terrorist attacks.
28. Manufacturing Company Employee Network (MCE; Cross *et al.*, 2004) [38]: This is a network of 77 employees (nodes) from a research team in a manufacturing company; there exists an edge between two nodes if the two employees are aware of each other's knowledge and skills.
29. Social Networks Journal Co-authors (MSJ; McCarty & Freeman, 2008) [39]: This is a network of 475 authors (vertices) involved in the production of 295 articles for the Social Networks Journal since its inception until 2008; there is an edge between two vertices if the corresponding authors co-authored at least one paper published in the journal.
30. Author Facebook Network (AFB): This is a network of the 171 friends (vertices) of the author in Facebook. There exists an edge between two vertices if the corresponding people are also friends of each other.
31. Mexican Political Elite Network (MPN; Gil-Mendieta & Schmidt, 1996) [40]: This is a network of 35 Mexican presidents and their close collaborators (vertices); there exists an edge between two vertices if the corresponding two people have ties that could be either political, kinship, friendship or business ties.
32. ModMath Network (MMN; Batagelj & Mrvar, 2006) [26]: This is a network of 30 school superintendents (vertices) in Allegheny County, Pennsylvania, USA during the 1950s and early 1960s. There exists an edge between two vertices if at least one of the two corresponding superintendents has indicated the other person as a friend in a research survey conducted to see which superintendents (who have been in office for at least a year) are more influential to effectively spread around some modern Math methods among the school systems in the county.
33. Network Science Co-authorship (NSC; Newman, 2006) [17]: This is a co-authorship network of 1589 scientists (vertices) working on network theory and experiments. There exists an edge between two vertices if the corresponding scientists have co-authored at least one publication in this area.
34. US Politics Books Network (PBN; Krebs, 2003) [41]: This is a network of 105 books (vertices) about US politics sold by Amazon.com around the time of the 2004 US presidential election. There exists an edge between two vertices if the corresponding two books were co-purchased by the same buyer (at least one buyer).
35. Primary School Contact Network (PSN; Gemmetto *et al.*, 2014) [42]: This is a network of children and teachers (238 vertices) used in the study published by an article in BMC Infectious Diseases, 2014 [40]. There exists an edge between two vertices if the corresponding persons were in contact for at least 20 seconds during the observation period.
36. Prison Friendship Network (PFN; MacRae, 1960) [43]: This is a network of 67 prison inmates (vertices) surveyed by John Gagnon in the 1950s regarding their sociometric choice. There exists an edge between two vertices if an inmate corresponding to at least one of them has listed the inmate corresponding to the other vertex as one of his / her closest friends.

37. San Juan Sur Family Network (SJN; Loomis *et al.*, 1953) [44]: This is a network of 75 families (vertices) in San Juan Sur, Costa Rica, 1948. There exists an edge between two vertices if at least one of the two families has visited the household of the family corresponding to the other vertex once or more.

38. Scotland Corporate Interlocks Network (SDI; Scott, 1980) [45]: This is a network of multiple directors (a director who serves on multiple boards) and companies (a total of 230 vertices) during 1904 – 05 in Scotland. There exists an edge between two vertices v_i and v_j if any of the following is true: (i) both v_i and v_j correspond to two different multiple directors who are in the board of at least one company; (ii) one of the two vertices corresponds to a multiple director and the other vertex corresponds to one of the companies in whose board the person serves.

39. Senator Press Release Network (SPR; Grimmer, 2010) [46]: This is a network of 92 US senators (vertices) during the period from 2007 to 2010. There exists an edge between two senators if they issued at least one joint press release.

40. Soccer World Cup 1998 Network (SWC; Batagelj & Mrvar, 2006) [26]: This is a network of 35 teams (vertices) that participated in the 1998 edition of the Soccer World Cup. A player for a national team could sometimes have contract with one or more other countries. In this network, there is an edge between two vertices if the national team corresponding to at least one of them has contracted players from the country represented by the national team corresponding to the other vertex.

41. Sawmill Strike Communication Network (SSM; Michael, 1997) [47]: This is a network of 24 employees (vertices) in a sawmill who planned a strike against the new compensation package proposed by their management. There exists an edge between any two vertices if the corresponding employees mutually admitted discussing about the strike with a frequency of three or more times during an observation period (on a 5-point scale).

42. Taro Exchange Network (TEN; Schwimmer, 1973) [48]: This is a network of 22 families (vertices) in a Papuan village. There exists an edge between two vertices if the corresponding families were seen exchanging gifts during an observation period.

43. Teenage Female Friendship Network (TWF; Pearson & Michell, 2000) [49]: This is a network of 47 female teenage students (vertices) who studied as a cohort in a school in the West of Scotland from 1995 to 1997. There exists an edge between two vertices if the corresponding students reported (in a survey) that they were best friends of each other.

44. UK Faculty Friendship Network (UKF; Nepusz *et al.*, 2008) [50]: This is a network of 83 faculties (vertices) at a UK university. There exists an edge between two vertices if the corresponding faculties are friends of each other.

45. US Airports 1997 Network (APN; Batagelj & Mrvar, 2006) [26]: This is a network of 332 airports (vertices) in the US in the year 1997. There is an edge between two nodes if there is a direct flight connection between the corresponding airports.

46. US States Network (USS) [54]: This is a network of the 48 contiguous states in the US and the District of Columbia (DC). Each of the 48 states and DC is a node and there is an edge involving two nodes if the corresponding states (or DC) have a common border between them.

47. Residence Hall Friendship Network (RHF; Freeman *et al.*, 1998) [51]: This is a network of 217 residents (vertices) living at a residence hall located on the Australian National University campus. There exists an edge between two vertices if the corresponding residents are friends of each other.

48. Windsurfers Beach Network (WSB; Freeman *et al.*, 1989) [52]: This is a network of 43 windsurfers (vertices) on a beach in southern California during Fall 1986. There exists an edge between two vertices if the corresponding windsurfers were perceived to be close to each other (determined via a survey).

49. World Trade Metal Network (WTN; Smith & White, 1992) [53]: This is a network of 80 countries (vertices) that were involved in trading miscellaneous metals during the period from 1965 to 1980. There exists an edge between two vertices if one of the two corresponding countries imported miscellaneous metals from the country corresponding to the other vertex.
50. Yeast Protein-Protein Interaction Network (YPI; Jeong *et al.*, 2001) [55]: This is a network of 1870 proteins (vertices) in Yeast and their mutual interactions modeled as undirected edges to study the correlation between lethality and centrality in protein-protein interaction (PPI) networks.

We measure the computation time per node (total computation time divided by the number of nodes) incurred for each of the five centrality metrics for the 50 real-world network graphs. The executions were conducted on a computer with Intel Core i7-2620M CPU @ 2.70 GHz and an installed main memory (RAM) of 8 GB. We ran the procedures for each of the five centrality metrics on each of the 50 real-world networks for 25 iterations and averaged the results. Table 2 lists the average computation time per node for the centrality metrics. For the computer architecture mentioned above and for the purpose of classification (as computationally-light vs. computationally-heavy centrality metrics), we consider a centrality metric as computationally-heavy if its average computation time per node is 0.01 millisecond or above (bold in Table 2) for at least 50% of the real-world networks analyzed, provided the suite of real-world networks analyzed is as diverse as it is in this paper (with respect to the number of nodes and edges and the fundamental metrics listed in Table 1). We observe the degree centrality metric to be computationally-light for all the real-world networks and the LCC'DC metric to be computationally-heavy for only 6% of the real-world networks. Hence, we refer to the DegC and LCC'DC metrics as computationally-light centrality metrics. On the other hand, we observe the CLC, EVC and BWC metrics to be computationally-heavy for 52%, 84% and 100% of the 50 real-world networks studied. Hence, we consider these three centrality metrics as computationally-heavy.

5. Correlation Analysis

In this section, we present in detail the results of the correlation analysis conducted for the computationally-light (DegC, LCC'DC) vs. computationally-heavy (CLC, EVC and BWC) centrality metrics (6 combinations of metrics) for the 50 real-world network graphs listed in Section 4. In order to validate our hypothesis, we measure the following:

(i) the difference in the correlation coefficient values between any two correlation measures;

(ii) the fraction (a total of $50 \cdot 6 = 300$ combinations) of the 50 real-world networks and the 6 combinations of computationally-light vs. computationally-heavy centrality metrics for which each of the three correlation measures incur the lowest correlation coefficient values;

(iii) the median of the correlation coefficient values observed for each correlation level between a computationally-light metric and a computationally-heavy metric.

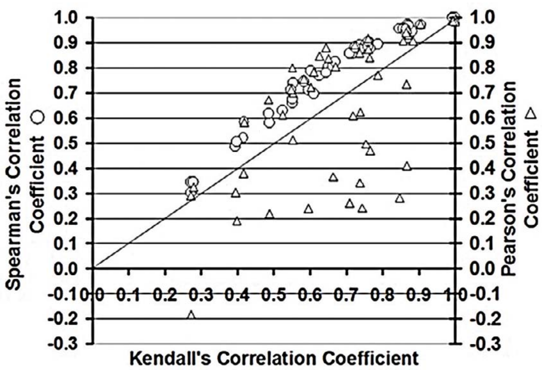
Table 3 lists the values for the correlation coefficient incurred with the three correlation measures (Kendall – τ; Spearman – ρ; Pearson – r) for DegC vs. {CLC, EVC and BWC} and Table 4 lists the values for LCC'DC vs. {CLC, EVC and BWC}. The correlation coefficient values reported in Tables 3 and 4 correspond to each of the 300 combinations of computationally-light vs. computationally-heavy centrality metrics and the real-world networks. In both these tables, we represent the cells (combinations) for which a correlation measure incurs the lowest value (in bold) and the largest value (in italics) for the correlation coefficient. We also plot the coefficient values (Kendall's vs. Spearman's and Pearson's correlation measures) in Figures 10 – 12 as well as the difference in the correlation coefficient values between any two correlation measures (in Figures 13 – 14) to facilitate visual analytics of the results.

5.1. On the Sufficiency of a Single Correlation Measure

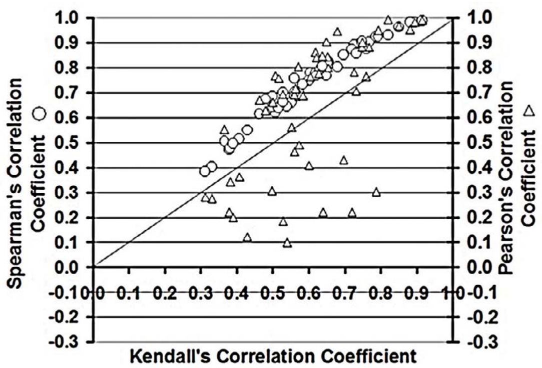
The results presented in Tables 3 – 4 and Figures 10 – 12 confirm our claim that a single correlation measure (like the most commonly used

Pearson's correlation measure) is not sufficient to assess all three levels of correlation. There is a total of 600 data points in Figures 10 – 12: if a single correlation measure is sufficient to assess all three levels of correlation, we would need a majority of these data points to fall on the diagonal line, implying that the correlation coefficient values determined using the three correlation measures should be equal or close enough to each other. However, we do not observe such a trend in Figures 10 – 12 as well as in Tables 3 – 4. There are several instances in Figures 10 – 12 for which the values for the Kendall's concordance-based correlation coefficient is significantly different from that of the Spearman's and Pearson's correlation coefficients. Since the Kendall's correlation coefficient is the most appropriate measure for the pair-wise concordance between two centrality metrics and the correlation coefficient values obtained with the Spearman's and Pearson's measures are appreciably different from the Kendall's correlation coefficient values, we conclude that the Spearman's and Pearson's measures cannot be used to assess the pair-wise concordance between a computationally-light metric in lieu of a computationally-heavy metric.

For each combination of computationally-light vs. computationally-heavy centrality metrics, we also determine the absolute value of the difference in the correlation coefficient values between any two correlation measures (i.e., Kendall's-Spearman's, Kendall's-Pearson's, and Spearman's-Pearson's) for the 50 real-world networks (resulting in a total of 150 data points for each of the six combinations of the centrality metrics) and plot a sorted list of the absolute difference in the correlation coefficient values, as shown in Figure 13. If all three correlation measures were to yield the same or close enough values for the correlation coefficient, we should have only obtained a flat line for each of the plots in Figure 13. However, we see that the difference in the correlation coefficient values could be as large as 0.3 – 0.7. We used threshold values of 0.05 and 0.10 for the difference in the correlation coefficient values and determined the fraction of the 150 data points (for each of the six combinations) for which the difference exceeds the threshold (see Figure 14). We observe that at least 40% of the data points had a difference in the correlation coefficient values of 0.10 or above for each of the six combinations of the centrality metrics



(a) DegC – CLC correlation analysis



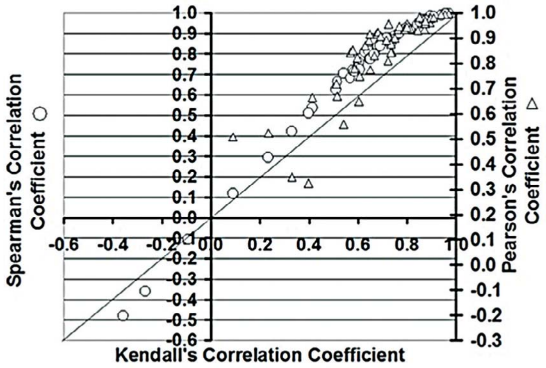
(b) LCC'DC – CLC correlation analysis

Figure 10. Kendall's vs. Spearman's and Pearson's correlation coefficients: {DegC, LCC'DC} vs. CLC.

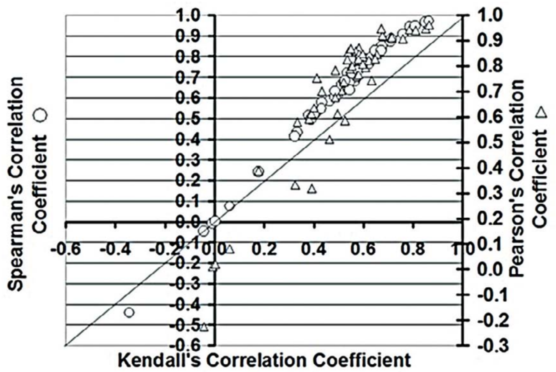
evaluated using any two of the three correlation measures. All of the above confirm that a single correlation measure would not be sufficient to assess all three levels of correlation.

5.2. Kendall's Correlation Measure for Lower Bound of the Correlation Coefficient

We observe the Kendall's correlation coefficient measure to incur the lowest of the correlation coefficient values for 114 of the 150 combinations in the case of DegC vs. the computationally-heavy centrality metrics {CLC, EVC and BWC} and for 111 of the 150 combinations in the case of BWC vs. the three computationally-heavy centrality metrics. Hence, we observe the Kendall's concordance-based correlation measure to be the lowest of the three correlation coefficient values for a total of (114 + 111) = 225 of the 300 combinations. As

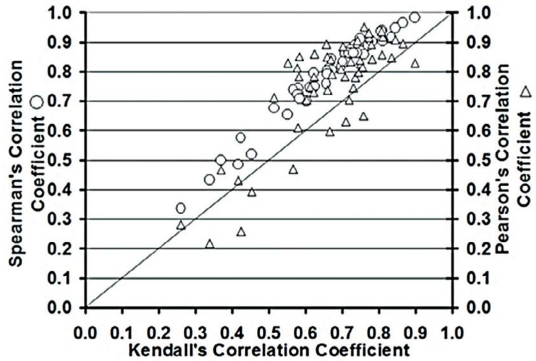


(a) DegC – EVC correlation analysis

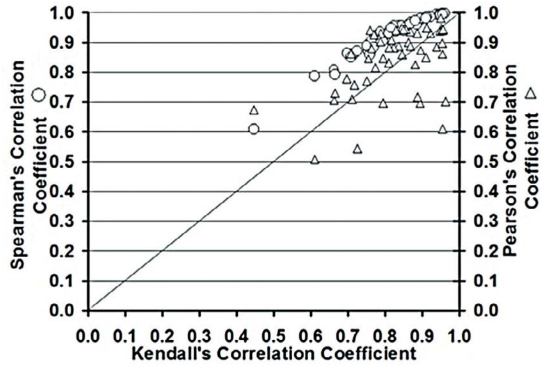


(b) LCC'DC – EVC correlation analysis

Figure 11. Kendall's vs. Spearman's and Pearson's correlation coefficients: {DegC, LCC'DC} vs. EVC.



(a) DegC – BWC correlation analysis



(b) LCC'DC – BWC correlation analysis

Figure 12. Kendall's vs. Spearman's and Pearson's correlation coefficients: {DegC, LCC'DC} vs. BWC.

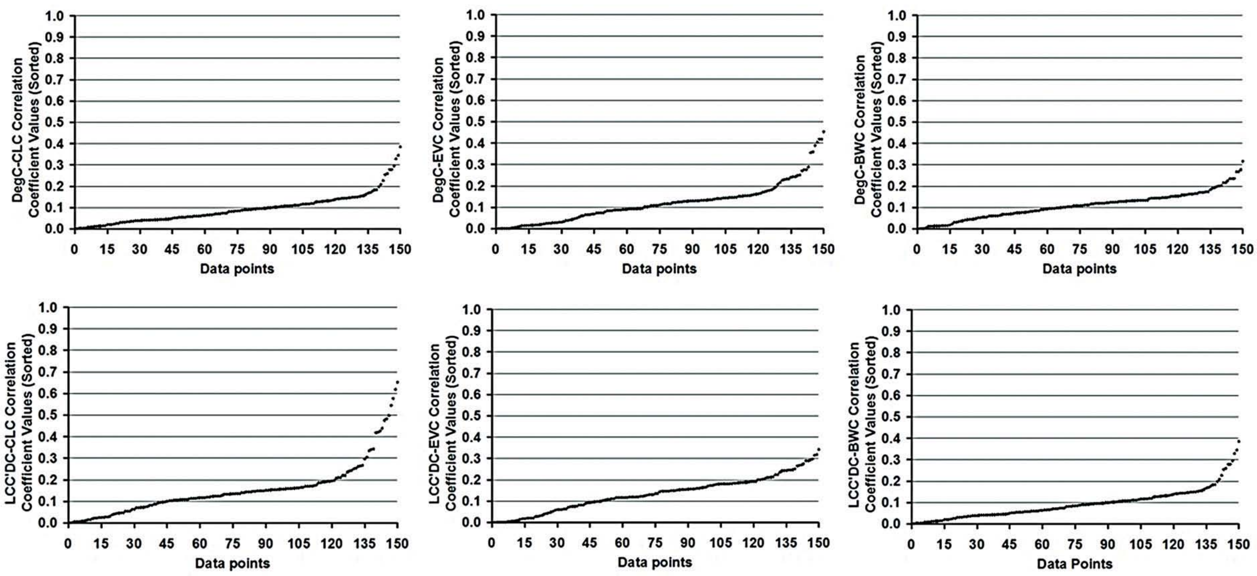


Figure 13. Distribution of the difference in the correlation coefficient values (sorted order) between any two correlation measures computed for {DegC, LCC'DC} vs. {CLC, EVC, BWC} metrics.

we analyze real-world networks whose degree distribution ranges from Poisson to Power-law (with spectral radius ratio for node degree [12] ranging from 1.01 to 5.5) and covering different domains (like social networks, citation networks, geographical networks, co-appearance networks, biological networks, etc), we claim that the 75% (or the equivalent decimal value of 225 / 300 = 0.75) could be considered as the probability with which the Kendall's concordance-based correlation coefficient observed for a computationally-light metric vs. a computationally-heavy metric would serve as a lower bound for the correlation coefficient expected between the same two centrality metrics under the Spearman's and Pearson's measures for any real-world network. The Spearman's rank-based correlation measure did not incur the lowest among the three correlation coefficient values for even one of the 300 combinations. The Pearson's correlation measure incurred the lowest correlation coefficient values for the remaining 25% of the 300 combinations of the computationally-light vs. computationally-heavy centrality metrics and the real-world network graphs.

Figures 10 – 12 present a visual analysis of the Kendall's correlation coefficient values vs. the Spearman's and Pearson's correlation coefficient values obtained for the computationally-light {DegC, LCC'DC} vs. the computationally-heavy {CLC, EVC and BWC} centrality metrics. If a data point lies above the diagonal line, then the Kendall's correlation coefficient for that combination is lower compared to the other correlation measure (either Spearman's or Pearson's, depending on the case). Hence, the larger the number of data points that are above the diagonal line, the larger the number of combinations of centrality metrics and real-world network graphs for which the Kendall's correlation coefficient is the lowest. We observe more than 95% of the blue data points (corresponding to the Spearman's correlation measure) to be above the diagonal line in both the sub-figures (a) and (b) of Figures 10 – 12. It is only the 25% of the red data points (corresponding to the Pearson's correlation measure) that are below the diagonal line, especially in the case of the computationally-light metrics vs. the CLC centrality metric. The Kendall's correlation coefficient is the lowest of the three correlation measures for more than 90% of the data points corresponding to the case of the computationally-light metrics vs. the EVC centrality metric.

The results thus convince us that the Kendall's concordance-based correlation measure should ideally be the first correlation measure one should compute between two centrality metrics (especially for correlation studies involving computationally-light vs. computationally-heavy centrality metrics) for a chosen real-world network and one should decide to proceed further based on the correlation coefficient value obtained. If we observe a strong correlation between a computationally-light centrality metric and a computationally-heavy centrality metric for a real-world network with respect to the Kendall's measure, there would not be even a need to compute the correlation coefficient with respect to the other two correlation measures (Spearman's and Pearson's) as there is a 0.75 chance that these correlation coefficient values will be at least the value observed for the Kendall's concordance-based correlation coefficient. From Tables 3 – 4, we also observe that the Kendall's correlation measure incurs the largest correlation coefficient value for only 8 of the 300 combinations (i.e., less than 3% of the 300 combinations). Hence, we have a strong implication that the probability that the Kendall's correlation coefficient is the largest of the three correlation values is as small as 3%.

5.3. Analysis of the Median Values for the Correlation Coefficient

Figures 15 (a) – (c) display the median values for the correlation coefficient observed for each of the three levels of correlation between a computationally-light metric {DegC, LCC'DC} with each of the computationally-heavy metrics {CLC, EVC, BWC} for the 50 real-world network graphs. Figure 15 (d) displays the median value when the correlation coefficient values for all three levels of correlation are considered together (hereafter referred to simply as *overall*) for a particular combination of the computationally-light and computationally-heavy metrics. Similar to the trend observed in Figures 10 – 12 and Tables 3 – 4, we also notice that irrespective of the computationally-light vs. computationally-heavy centrality metric combination, the median value of the correlation coefficient observed for pair-wise relative ordering of the vertices is the lowest among the

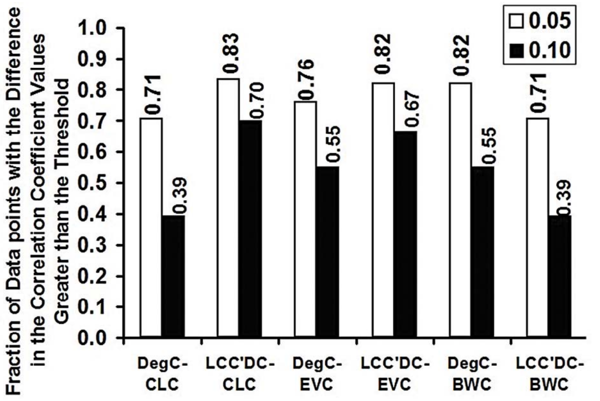
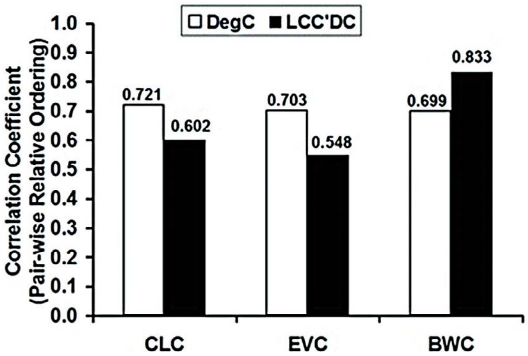


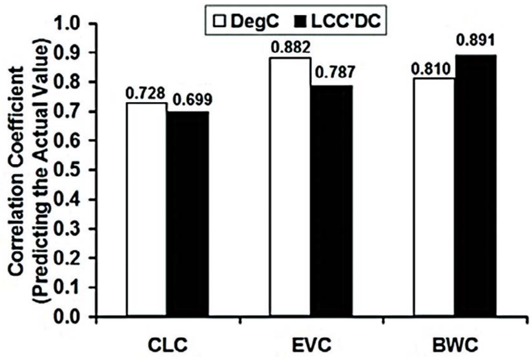
Figure 14. Fraction of the data points with the difference in the correlation coefficient values between any two correlation measures greater than threshold values of 0.05 and 0.10.

correlation coefficient values for all three levels of correlation.

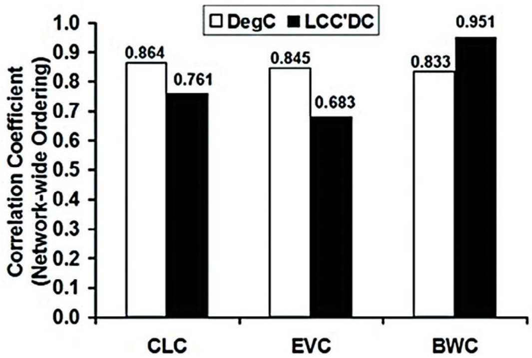
With respect to the individual combination of centrality metrics: we consistently observe DegC to exhibit higher levels of correlation with the closeness and eigenvector centrality metrics for each of the three levels of correlation as well as for overall; whereas, we observe the local clustering coefficient complement-based degree centrality (LCC'DC) metric to exhibit relatively stronger correlation with the betweeness centrality (BWC) metric for each of the three levels of correlation as well as for overall. We thus conclude that for each of the three levels of correlation: the DegC metric could serve as the computationally-light alternative for the CLC and EVC metrics; whereas, LCC'DC could serve as the computationally-light alterative for BWC.



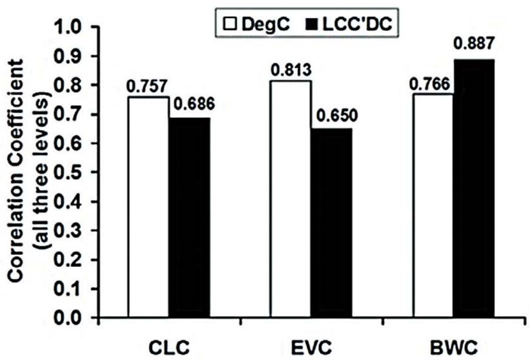
(a) Paire-wise relative ordering (Kendall's measure).



(c) Predicting thr actual values (Pearson's measure).



(b) Network-wide ranking (Spearman's measure).



(d) NAll three levels (Kendall's, Spearman's, Pearson's measure).

Figure 15. Median values for the correlation coefficient for each level of correlation and all three levels: {DegC, LCC'DC} vs. {CLC, EVC, BWC}.

Table 3. Degree centrality vs. computationally-heavy metrics: results of correlation analysis.

		Degree Centrality (DegC) vs. Closeness Centrality (CLC)			**Degree Centrality (DegC) vs. Eigenvector Centrality (EVC)**			**Degree Centrality (DegC) vs. Betweenness Centrality (BWC)**		
#	**Net.**	**Kendall**	**Spearman**	**Pearson**	**Kendall**	**Spearman**	**Pearson**	**Kendall**	**Spearman**	**Pearson**
1	ADJ	**0.764**	*0.901*	0.841	**0.801**	0.929	*0.957*	**0.773**	0.901	*0.915*
2	AKN	**0.626**	0.767	*0.846*	**0.763**	0.897	*0.936*	**0.657**	0.759	*0.892*
3	JBN	**0.736**	*0.890*	0.859	**0.750**	0.890	*0.901*	**0.579**	*0.744*	0.610
4	CEN	**0.553**	*0.738*	0.700	**0.629**	0.811	*0.871*	**0.736**	*0.889*	0.780
5	CLN	0.847	*0.956*	**0.282**	**0.892**	*0.976*	0.961	**0.750**	*0.903*	0.825
6	CGD	0.754	*0.893*	**0.497**	**0.722**	*0.876*	0.810	**0.745**	*0.890*	0.797
7	CFN	**0.882**	*0.945*	0.908	**0.870**	*0.965*	0.935	0.697	*0.818*	0.808
8	DON	**0.548**	*0.718*	0.713	**0.512**	0.627	*0.720*	0.667	*0.814*	**0.598**
9	DRN	0.718	*0.856*	**0.608**	**0.603**	*0.758*	0.650	0.758	*0.875*	**0.649**
10	DLN	0.856	*0.953*	0.908	**0.768**	0.904	*0.947*	**0.672**	*0.804*	0.791
11	ERD	0.709	*0.858*	**0.261**	**0.675**	0.827	*0.916*	**0.708**	*0.860*	0.782
12	FMH	0.739	*0.871*	**0.624**	**0.541**	*0.704*	0.558	0.711	*0.832*	**0.630**
13	FHT	0.866	*0.956*	**0.409**	**0.812**	0.920	*0.937*	**0.755**	*0.902*	0.816
14	FTC	**0.650**	0.802	*0.837*	**0.596**	0.730	*0.822*	**0.582**	0.723	*0.783*
15	FON	**0.272**	*0.344*	0.291	**0.606**	0.722	*0.750*	**0.260**	*0.336*	0.282
16	CDF	0.998	*1.000*	**0.990**	**0.972**	0.991	*0.997*	**0.809**	*0.940*	0.857
17	GD96	0.552	*0.659*	**0.513**	**0.568**	0.684	*0.844*	**0.759**	0.859	*0.951*
18	MUN	0.395	*0.486*	**0.303**	*-0.356*	-0.479	**-0.712**	**0.603**	0.699	*0.704*
19	GLN	0.664	*0.806*	**0.366**	**0.578**	0.718	*0.853*	**0.773**	0.888	*0.932*
20	HTN	**0.990**	*0.999*	0.993	**0.954**	*0.995*	0.994	0.899	*0.983*	**0.829**
21	HCN	0.743	*0.874*	**0.241**	**0.791**	0.922	*0.936*	**0.552**	0.656	0.829
22	ISP	**0.602**	*0.786*	0.722	**0.644**	0.813	*0.893*	0.566	0.737	**0.469**
23	KCN	0.786	*0.895*	**0.772**	**0.647**	0.775	*0.917*	**0.811**	0.905	*0.918*
24	KFP	0.766	*0.877*	**0.470**	**0.843**	*0.945*	0.931	**0.370**	*0.500*	0.467
25	LMN	**0.551**	0.675	*0.800*	**0.738**	*0.868*	0.847	**0.612**	0.745	*0.747*
26	MDN	0.997	1.000	**0.992**	**0.940**	0.990	*0.994*	**0.807**	*0.936*	0.935
27	MTB	0.737	0.872	**0.341**	**0.682**	0.835	*0.924*	**0.622**	*0.746*	0.729
28	MCE	0.990	*1.000*	**0.982**	**0.874**	0.957	*0.977*	**0.701**	0.834	*0.885*
29	MSJ	0.488	*0.580*	**0.217**	**0.090**	0.120	*0.508*	0.453	*0.520*	**0.392**
30	AFB	0.272	*0.303*	**-0.183**	*-0.267*	-0.361	**-0.720**	0.424	*0.576*	**0.259**
31	MPN	**0.643**	0.780	*0.881*	**0.692**	0.838	0.907	**0.780**	*0.905*	0.892
32	MMN	0.865	*0.943*	**0.733**	**0.734**	0.851	0.877	**0.781**	*0.903*	0.842
33	NSC	0.595	*0.711*	**0.240**	*-0.092*	-0.107	**-0.511**	**0.416**	*0.485*	0.431
34	PBN	**0.418**	*0.585*	0.582	**0.515**	0.663	*0.670*	**0.515**	0.677	*0.712*
35	PSN	**0.869**	*0.974*	0.952	**0.895**	*0.983*	0.982	**0.749**	*0.913*	0.838
36	PFN	**0.761**	*0.884*	0.875	**0.733**	*0.863*	0.843	**0.659**	0.804	*0.849*
37	SJN	**0.486**	0.618	*0.672*	**0.413**	0.536	*0.664*	**0.577**	0.722	*0.812*
38	SDI	0.416	*0.520*	**0.379**	0.398	*0.512*	**0.324**	**0.660**	*0.792*	0.737
39	SPR	**0.870**	*0.968*	0.930	**0.866**	0.968	*0.976*	*0.723*	*0.872*	0.835
40	SWC	**0.864**	*0.954*	0.941	**0.874**	0.964	*0.968*	*0.742*	0.863	*0.905*
41	SSM	**0.610**	0.696	*0.782*	**0.585**	0.714	*0.780*	*0.584*	0.708	*0.851*
42	TEN	**0.524**	*0.629*	0.612	**0.650**	0.774	*0.776*	*0.624*	0.750	*0.859*
43	TWF	**0.279**	*0.344*	0.326	**0.235**	0.294	*0.523*	0.338	*0.433*	**0.218**
44	UKF	**0.759**	0.904	*0.918*	**0.799**	0.928	*0.944*	**0.624**	*0.794*	0.782
45	APN	**0.670**	*0.823*	0.803	**0.725**	0.864	*0.956*	0.719	*0.863*	**0.705**
46	USS	**0.582**	0.746	*0.755*	**0.667**	0.799	*0.832*	**0.730**	*0.864*	0.744
47	RHF	**0.724**	0.881	*0.891*	**0.715**	0.876	*0.892*	**0.669**	*0.843*	0.841
48	WSB	**0.904**	0.971	*0.975*	**0.909**	0.983	0.982	**0.866**	*0.964*	0.895
49	WTN	0.993	*0.999*	**0.987**	**0.851**	0.954	*0.983*	**0.845**	*0.949*	0.908
50	YPI	0.398	*0.506*	**0.191**	**0.330**	*0.422*	0.349	**0.834**	*0.917*	0.847
# Lowest		**27/50**	**23/50**	**23/50**	**46/50**	**0/50**	**4/50**	**41/50**	**0**	**9/50**
# Largest		*0/50*	*40/50*	*10/50*	*3/50*	*13/50*	*34/50*	*4/50*	*31/50*	*15/50*

Table 4. LCC'DC vs. computationally-heavy metrics: results of correlation analysis.

		LCC'DC vs. Closeness Centrality (CLC)			**LCC'DC vs. Eigenvector Centrality (EVC)**			**LCC'DC vs. Betweenness Centrality (BWC)**		
#	**Net.**	**Kendall**	**Spearman**	**Pearson**	**Kendall**	**Spearman**	**Pearson**	**Kendall**	**Spearman**	**Pearson**
1	ADJ	**0.655**	*0.824*	0.799	**0.676**	0.850	*0.920*	**0.789**	0.916	*0.930*
2	AKN	**0.507**	0.621	*0.769*	**0.540**	0.664	*0.855*	0.951	*0.994*	**0.948**
3	JBN	**0.726**	*0.891*	0.782	**0.608**	0.788	*0.793*	**0.717**	*0.860*	0.757
4	CEN	**0.499**	*0.685*	0.661	**0.535**	0.719	*0.825*	**0.774**	*0.923*	0.816
5	CLN	0.720	*0.874*	0.221	**0.759**	*0.908*	0.907	**0.837**	*0.954*	0.887
6	CGD	0.697	*0.852*	**0.432**	**0.633**	*0.797*	0.744	**0.846**	*0.956*	0.860
7	CFN	**0.649**	0.767	0.903	**0.622**	0.766	*0.823*	0.954	*0.993*	**0.897**
8	DON	**0.604**	*0.780*	0.765	**0.513**	0.663	*0.703*	0.711	*0.861*	**0.709**
9	DRN	0.573	*0.700*	**0.490**	**0.495**	0.610	*0.613*	0.894	*0.975*	**0.696**
10	DLN	**0.768**	*0.903*	0.882	**0.654**	0.817	*0.845*	**0.755**	*0.872*	0.846
11	ERD	0.639	*0.798*	**0.221**	**0.581**	0.741	*0.870*	**0.810**	*0.936*	0.831
12	FMH	0.560	*0.684*	**0.464**	**0.463**	*0.586*	0.511	0.888	*0.973*	**0.718**
13	FHT	0.787	*0.923*	**0.303**	**0.646**	0.827	*0.829*	**0.863**	*0.959*	0.900
14	FTC	**0.652**	0.821	*0.845*	**0.432**	0.579	*0.700*	**0.784**	*0.918*	0.913
15	FON	**0.366**	0.506	*0.552*	**-0.009**	-0.007	*0.011*	**0.447**	0.608	*0.673*
16	CDF	**0.895**	0.981	*0.982*	**0.850**	*0.967*	0.946	**0.869**	*0.968*	0.935
17	GD96	**0.552**	*0.659*	0.562	**0.568**	0.684	*0.860*	**0.759**	0.859	*0.942*
18	MUN	0.379	*0.472*	**0.222**	*-0.344*	-0.440	**-0.548**	0.955	*0.995*	**0.861**
19	GLN	0.498	*0.642*	**0.307**	**0.411**	0.530	*0.753*	**0.856**	0.952	0.944
20	HTN	**0.914**	0.987	*0.990*	**0.864**	*0.972*	0.963	0.939	*0.994*	**0.884**
21	HCN	0.539	*0.645*	**0.100**	**0.486**	0.603	*0.784*	0.948	*0.993*	**0.938**
22	ISP	**0.559**	*0.756*	0.692	**0.583**	0.771	*0.848*	0.611	*0.787*	**0.509**
23	KCN	**0.759**	*0.874*	0.766	**0.549**	0.680	*0.867*	**0.886**	*0.960*	0.930
24	KFP	0.600	*0.749*	**0.408**	**0.521**	0.674	*0.736*	**0.663**	*0.807*	0.705
25	LMN	**0.516**	0.639	*0.757*	**0.525**	*0.683*	0.585	**0.923**	*0.987*	0.931
26	MDN	**0.792**	0.925	*0.950*	**0.711**	0.871	*0.913*	**0.950**	*0.995*	0.982
27	MTB	0.528	*0.662*	**0.186**	**0.431**	0.548	*0.701*	0.896	*0.981*	**0.874**
28	MCE	**0.679**	0.802	*0.946*	**0.546**	0.638	*0.790*	0.955	*0.996*	**0.942**
29	MSJ	0.331	*0.401*	**0.277**	**0.061**	0.076	*0.082*	0.955	*0.996*	**0.610**
30	AFB	0.429	*0.549*	**0.123**	-0.044	-0.045	**-0.224**	0.726	*0.871*	**0.543**
31	MPN	**0.618**	0.780	*0.862*	**0.597**	0.778	*0.838*	**0.830**	0.938	*0.941*
32	MMN	0.732	*0.856*	**0.705**	**0.573**	0.701	*0.761*	**0.868**	*0.962*	0.888
33	NSC	0.312	*0.383*	**0.281**	**0.003**	0.005	*0.020*	0.963	*0.997*	**0.703**
34	PBN	**0.479**	*0.674*	0.627	**0.381**	0.516	*0.591*	**0.691**	*0.864*	0.779
35	PSN	**0.881**	*0.981*	0.954	**0.786**	0.941	*0.943*	**0.824**	*0.955*	0.883
36	PFN	**0.628**	*0.788*	0.777	**0.485**	0.632	*0.677*	**0.811**	*0.929*	0.882
37	SJN	**0.462**	0.615	*0.670*	**0.333**	0.432	*0.579*	**0.708**	0.851	*0.861*
38	SDI	0.407	*0.514*	**0.363**	0.391	0.502	**0.318**	**0.665**	*0.793*	0.730
39	SPR	**0.747**	*0.904*	0.882	**0.713**	0.886	*0.914*	**0.763**	*0.905*	0.880
40	SWC	**0.621**	0.768	*0.841*	**0.597**	0.767	*0.848*	**0.771**	0.883	*0.927*
41	SSM	**0.570**	0.686	*0.804*	**0.390**	0.494	*0.613*	**0.795**	*0.906*	0.847
42	TEN	**0.562**	0.703	*0.717*	**0.401**	0.520	*0.636*	**0.850**	0.939	*0.942*
43	TWF	0.382	*0.478*	**0.344**	**0.177**	0.241	*0.388*	0.795	*0.904*	**0.696**
44	UKF	**0.637**	0.806	*0.848*	**0.554**	0.719	*0.801*	**0.818**	*0.949*	0.908
45	APN	**0.583**	*0.733*	0.687	**0.579**	0.735	*0.827*	0.882	*0.973*	**0.825**
46	USS	**0.528**	*0.701*	0.693	**0.579**	0.733	*0.766*	**0.751**	*0.889*	0.770
47	RHF	**0.748**	*0.907*	0.902	**0.596**	0.777	*0.808*	**0.787**	*0.934*	0.903
48	WSB	**0.850**	0.962	*0.967*	**0.810**	*0.947*	0.940	**0.912**	*0.986*	0.948
49	WTN	**0.820**	0.929	*0.992*	**0.672**	0.827	*0.948*	0.956	*0.995*	**0.944**
50	YPI	0.390	*0.496*	**0.199**	**0.324**	*0.414*	0.333	0.910	*0.980*	**0.849**
# Lowest		**32/50**	**0/50**	**18/50**	**47/50**	**0/50**	**3/50**	**32/50**	**0/50**	**18/50**
# Largest		*0/50*	*33/50*	*17/50*	*1/50*	*8/50*	*41/50*	*0/50*	*43/50*	*7/50*

5.4. LCC'DC-BWC Correlation vs. DegC-BWC Correlation

We observe the LCC'DC metric to exhibit a very strong correlation with the BWC metric (the most time-consuming metric of the three computationally-heavy centrality metrics) and the data points (in Figure 12 (b), with respect to the three correlation measures, are located relatively closer to the largest possible value of 1 and also closer to each other for a majority of the real-world network graphs. Considering a total of 150 LCC'DC-BWC correlation coefficient values obtained with respect to the three correlation measures for the 50 real-world network graphs, we observe (see Figure 15 (d)) the median value to be 0.887 (the largest median value for each of the six combinations of computationally-light vs. computationally-heavy centrality metrics: see Figures 15 (a), (b) and (c) and only 12 of the 150 correlation coefficient values (i.e., less than 10%) are below 0.7. On the other hand, the median value of the 150 DegC-BWC correlation coefficient values for the three correlation measures analyzed for the 50 real-world network graphs is 0.766 (see Figure 15 (d)), appreciably lower than the median value of 0.887 for the LCC'DC-BWC correlation.

6. Related Work and Our Contributions

The idea of studying correlation between computationally-light centrality metrics and the computationally-heavy centrality metrics was recently mooted by Li *et al.* [5] in which the Pearson's correlation coefficient was used as the correlation measure to evaluate the extent to which one could rank the vertices using a computationally-light centrality metric in lieu of a computationally-heavy centrality metric. However, as seen in this paper, the Pearson's correlation coefficient values are different from those of the Spearman's and Kendall's rank-based correlation measures for at least the computationally-light centrality metrics vs. computationally-heavy shortest path-based centrality metrics. In another recent work [56], it has been observed that the Kendall's concordance-based correlation measure is more suitable to evaluate pair-wise correlation, especially among the top-*k* ranked vertices, whereas the Spearman's correlation measure is more suitable to evaluate rank-based correlation involving all the vertices, especially if several of them have equal ranks. The three correlation measures were also recently used [57] to analyze the extent to which one can predict flux changes using the functional centrality metric [58] proposed to quantify the functional relevance of individual biochemical reactions in metabolic networks. In [8], the computationally-light degree centrality metric and the computationally-heavy eigenvector centrality metric were observed to be strongly correlated with the computationally-heavy maximal clique size for a node (the maximal clique size for a node is the clique of the largest size that a node is part of) under all three levels of correlation. The BWC metric has been observed to be weakly correlated (correlation coefficient values in the range 0...0.5) with the maximal clique size. Unlike the work in [8] wherein the correlation between centrality metrics and maximal clique size was studied, in this paper: we investigate the correlation among the centrality metrics themselves on the lines of computationally-light {DegC, LCC'DC} vs. computationally-heavy {CLC, EVC, BWC} centrality metrics. The LCC'DC metric was also not considered in [8].

In [59], the author developed a new centrality metric called CIRank (that keeps track of the changes propagating among classes in a software dependency network) and observed it to be significantly correlated with the degree and PageRank centrality metrics on the basis of the Spearman's rank-based correlation coefficient. In [60], the Kendall's concordance-based correlation measure was used to assess the correlation between eight different centrality metrics that are suitable for gene regulatory networks in E. Coli. It was observed that the ranking of the genes with respect to the centrality metrics is significantly different (leading to a low correlation coefficient between any two centrality metrics), especially when vertices (genes) with non -zero out degree are only considered. In another related study [61], the Kendall's measure was used to study the correlation between DegC, CLC, BWC and EVC metrics for the *M. musculus* protein-protein interaction network. In [62], the authors studied the impact of removing the top-k ranked vertices (with respect to a centrality metric) on the traffic-carrying capacity of the remaining nodes and the connectivity of ISP (Internet Service Provider) networks: Removal of the top-*k* vertices with respect to the locally computable degree centrality metric had a similar impact on the traffic-carrying capacity of the remaining nodes vis-a-vis removal of the top-*k* vertices with respect to the globally computable centrality metrics.

Though some of the recent works in the literature (as mentioned above) have also used the three correlation measures (Kendall's, Spearman's and Pearson's) for analyzing the correlation between centrality metrics with respect to the three levels of correlation (pair-wise relative ordering, network-wide ranking, and prediction of the actual values), ours is the first work to evaluate the three levels of correlation from the point of view of computationally-light vs. computationally-heavy centrality metrics and demonstrate that the pair-wise relative ordering of the vertices could be the most restrictive and that the corresponding Kendall's concordance-based correlation measure could serve (with a probability as large as 0.75) as the lower bound for correlation coefficient among the three levels of correlation. We could also conclude that the Kendall's correlation coefficient is not the largest among the correlation coefficients of the three measures with a probability of 0.97. Moreover, the results of the extensive correlation studies in this paper (conducted on a wide range of 50 real-world networks) also reaffirm the status of LCC'DC to be a computationally-light alternative that could be considered alongside the degree centrality metric and could be strongly correlated to one or more of the computationally-heavy metrics, especially the BWC metric with respect to all three levels of correlation.

7. Conclusion

We observe the pair-wise relative ordering of vertices based on a computationally-light metric in lieu of a computationally-heavy centrality metric to be the most restrictive of all three levels of correlation and the Kendall's concordance-based correlation coefficient (that is a measure of the level of correlation to assess the pair-wise relative ordering of vertices) could be considered (with a probability as large as 0.75) to serve as the lower bound for correlation coefficient between a computationally-light and computationally-heavy centrality metric. Likewise, we could also conclude that the Kendall's correlation coefficient is not the largest of the three correlation measures (between a computationally-light and computationally-heavy centrality metric) with a probability of 0.97. Such significant observations on the nature of the correlation coefficient values obtained for the centrality metrics (especially for computationally-light vs. computationally-heavy metrics) with respect to the Kendall's concordance-based correlation measure have been hitherto not reported in the literature. In addition, we also observe the correlation between LCC'DC and BWC to be the strongest of all the different combinations of the computationally-light and computationally-heavy centrality metrics for each of the three levels of correlation as well as when all three levels of correlation are considered together. Until now, the degree centrality metric has been so far considered the computationally-light alternative for any computationally-heavy centrality metric, including the BWC. Through this paper, we have established that LCC'DC is a relatively better computationally-light alternative to BWC with respect to all three levels of correlation, including the pair-wise relative ordering of the vertices.

References

[1] M. E. J. Newman, "Networks: An Introduction", 1st Edition, Oxford University Press, Oxford, UK, 2010.

[2] P. Bonacich, "Power and Centrality: A Family of Measures", *American Journal of Sociology*, vol. 92, no. 5, pp. 1170–1182, 1987. http://dx.doi.org/10.1086/228631

[3] L. Freeman, "A Set of Measures of Centrality based on Betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977. http://dx.doi.org/10.2307/3033543

[4] L. Freeman, "Centrality in Social Networks Conceptual Clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1979. http://dx.doi.org/10.1016/0378-8733(78)90021-7

[5] C. Li, Q. *et al.*, "Correlation between Centrality Metrics and their Application to the Opinion

Model," *The European Physical Journal B*, vol. 88, no. 65, pp. 1–13, 2015.
http://dx.doi.org/10.1140/epjb/e2015-50671-y

[6] N. Meghanathan, "Correlation Coefficient Analysis of Centrality Metrics for Complex Network Graphs," *Proceedings of the 4th Computer Science Online Conference, (CSOC-2015), Intelligent Systems in Cybernetics and Automation Theory: Advances in Intelligent Systems and Computing*, vol. 348, 2015, pp. 11–20.
http://dx.doi.org/ 10.1007/978-3-319-18503-3_2

[7] M. F. Triola, "Elementary Statistics", 12th Edition, Pearson, NY, USA, 2012.

[8] N. Meghanathan, "Maximal Clique Size vs. Centrality: A Correlation Analysis for Complex Real-World Network Graphs," *Proceedings of the 3rd International Conference on Advanced Computing, Networking, and Informatics, Springer Smart Innovation, Systems and Technologies Series*, Orissa, India, vol. 44, pp. 95–101.
http://dx.doi.org/10.1007/978-81-322-2529-4_9

[9] N. Meghanathan, "A Computationally Lightweight and Localized Centrality Metric in lieu of Betweenness Centrality for Complex Network Analysis," *Springer Vietnam Journal of Computer Science*, vol. 4, no. 1, pp. 23–38, 2017.
http://dx.doi.org/10.1007/s40595-016-0073-1

[10] N. Meghanathan and X. He, "Correlation and Regression Analysis for Node Betweenness Centrality," *International Journal of Foundations in Computer Science and Technology*, vol. 6, no. 6, pp. 1–20, 2016.
http://dx.doi.org/10.5121/ijfcst.2016.6601

[11] V. Krebs, "Proxy Networks: Analyzing One Network to Reveal Another," *Bulletin de Méthodologie Sociologique*, vol. 79, pp. 61–40, 003.
http://dx.doi.org/10.1177/075910630307900105

[12] N. Meghanathan, "Spectral Radius as a Measure of Variation in Node Degree for Complex Network Graphs," *Proceedings of the 7th International Conference on u- and e- Service, Science and Technology*, Haikou, China, 2014, pp. 30–33.
http://dx.doi.org/ 10.1109/UNESST.2014.8

[13] D. C. Lay, "Linear Algebra and its Applications", 4th Edition, Pearson, NY, USA, 2011.

[14] U. Brandes, "A Faster Algorithm for Betweenness Centrality," *The Journal of Mathematical Sociology*, vol. 25, no. 2, pp. 163–177, 2001.
http://dx.doi.org/10.1080/0022250X.2001.9990249

[15] T. H. Cormen *et al.*, "Introduction to Algorithms", 3rd Edition, MIT Press, Cambridge, MA, USA, 2009.

[16] M. E. J. Newman, "Modularity and Community Structure in Networks," *Proceedings of the National Academy of Sciences USA*, 2006, vol. 103, no. 23, pp. 8557–8582.
http://dx.doi.org/10.1073/pnas.0601602103

[17] M. E. J. Newman, "Finding Community Structure in Networks using the Eigenvectors of Matrices," *Physical Review E*, vol. 74, no. 3, 036104, 2006.
http://dx.doi.org/10.1103/PhysRevE.74.036104

[18] D. E. Knuth, "The Stanford GraphBase: A Platform for Combinatorial Computing", 1st Edition, Addison-Wesley, Reading, MA, 1993.

[19] P. Geiser and L. Danon, "Community Structure in Jazz," *Advances in Complex Systems*, vol. 6, no. 4, pp. 563–573, 2003.
http://dx.doi.org/10.1142/S0219525903001067

[20] J. G. White *et al.*, "The Structure of the Nervous System of the Nematode Caenorhabditis Elegans", *Philosophical Transactions B*, vol. 314, no. 1165, pp. 1–340, 1986.
http://dx.doi.org/10.1098/rstb.1986.0056

[21] N. P. Hummon *et al.*, "Analyzing the Structure of the Centrality-Productivity Literature Created between 1948 and 1979," *Science Communication*, vol. 11, no. 4, pp. 459–480, 1990.
http://dx.doi.org/10.1177/107554709001100405

[22] T. Biedl and B. J. Franz, "Graph-Drawing Contest Report," *Proceedings of the 9th International Symposium on Graph Drawing*, pp. 513–521, 2001.
http://dx.doi.org/10.1007/978-3-642-18469-7_44

[23] D. Lusseau *et al.*, "The Bottlenose Dolphin Community of Doubtful Sound Features a Large Proportion of Long-lasting Associations", *Behavioral Ecology and Sociobiology*, vol. 54, no. 3, pp. 396–405, 2003.
http://dx.doi.org/10.1007/s00265-003-0651-y

[24] J.-S. Lee, "Generating Networks of Illegal Drug Users Using Large Samples of Partial Ego-Network Data," *Intelligence and Security Informatics, Lecture Nodes in Computer Science*, vol. 3073, pp. 390–402, 2004.
http://dx.doi.org/10.1007/978-3-540-25952-7_29

[25] W. de Nooy, "A Literary Playground: Literary Criticism and Balance Theory," *Poetics*, vol. 26, no. 5–6, pp. 385–404, 1999.
http://dx.doi.org/10.1016/S0304-422X(99)00009-1

[26] V. Batagelj and A. Mrvar, Pajek Datasets, 2006.
http://vlado.fmf.uni-lj.si/pub/networks/data/

[27] M. D. Resnick *et al.*, "Protecting Adolescents from Harm. Findings from the National Longitudinal Study on Adolescent Health", *Journal of the American Medical Association*, vol. 278, no. 10, pp. 823–832, 1997.
http://dx.doi.org/10.1001/jama.1997.03550100049038

[28] D. Krackhardt, "The Ties that Torture: Simmelian Tie Analysis in Organizations," *Research in the Sociology of Organizations*, vol. 16, pp. 183-210, 1999.

[29] J. L. Moreno, "The Sociometry Reader", The Free Press, Glencoe, IL, USA, 1960, pp. 534–547.

[30] M. Girvan and M. E. J. Newman, "Community Structure in Social and Biological Networks", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7821–7826, 2002. http://dx.doi.org/10.1073/pnas.122653799

[31] H. R. Bernard *et al.*, "Informant Accuracy in Social Network Data IV: A Comparison of Clique-level Structure in Behavioral and Cognitive Network Data", *Social Networks*, vol. 2, no. 3, pp. 191–218, 1980. http://dx.doi.org/10.1016/0378-8733(79)90014-5

[32] P. M. Gleiser, "How to become a Superhero", *Journal of Statistical Mechanics: Theory and Experiments*, P09020, 2007.

[33] L. Isella *et al.*, "What's in a Crowd? Analysis of Face-to-Face Behavioral Networks", *Journal of Theoretical Biology*, vol. 271, no. 1, pp. 166–180, 2011. http://dx.doi.org/10.1016/j.jtbi.2010.11.033

[34] W. W. Zachary, "An Information Flow Model for Conflict and Fission in Small Groups", *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452–473, 1977.

[35] E. M. Rogers and D. L. Kincaid, "Communication Networks: Toward a New Paradigm for Research", Free Press, 1980.

[36] Y. Takahata, "Diachronic Changes in the Dominance Relations of Adult Female Japanese Monkeys of the Arashiyama B Group", The Monkeys of Arashiyama, Albany: State University of New York Press, 1991, pp. 124–139.

[37] B. Hayes, "Connecting the Dots", *American Scientist*, vol. 94, no. 5, pp. 400–404, 2006. http://dx.doi.org/10.1511/2006.61.3495

[38] R. L. Cross *et al.*, "The Hidden Power of Social Networks: Understanding How Work Really Gets Done in Organizations", Harvard Business Review Press, 1st Edition, 2004.

[39] C. McCarty and L. Freeman, 2008. http://moreno.ss.uci.edu/data.html

[40] J. Gil-Mendieta and S. Schmidt, "The Political Network in Mexico", *Social Networks*, vol. 18, no. 4, pp. 355–381, 1996. http://dx.doi.org/10.1016/0378-8733(95)00281-2

[41] V. Krebs, "Proxy Networks: Analyzing One Network to Reveal Another," *Bulletin de Méthodologie Sociologique*, vol. 79, pp. 61–40, 2003. http://dx.doi.org/10.1177/075910630307900105

[42] V. Gemmetto *et al.*, "Mitigation of Infectious Disease at School: Targeted Class Closure vs. School Closure". *BMC Infectious Diseases*, vol. 14, no. 695, pp. 1–10, 2014. http://dx.doi.org/10.1186/s12879-014-0695-9

[43] D. MacRae, "Direct Factor Analysis of Sociometric Data", *Sociometry*, vol. 23, no. 4, pp. 360–371, 1960.

[44] C. P. Loomis *et al.*, "Turrialba Social Systems and the Introduction of Change", The Free Press, Glencoe, IL, USA, 1953, pp. 45–78.

[45] J. P. Scott, "The Anatomy of Scottish Capital: Scottish Companies and Scottish Capital, 1900 – 1979", Croom Helm, 1st Edition, 1980.

[46] J. Grimmer, "A Bayesian Hierarchical Topic Mode for Political Texts: Measuring Expressed Agendas in Senate Press Releases", *Political Analysis*, vol. 18, no. 1, pp. 1–35, 2010. http://dx.doi.org/10.1093/pan/mpp034

[47] J. H. Michael, "Labor Dispute Reconciliation in a Forest Products Manufacturing Facility", *Forest Products Journal*, vol. 47, no. 11–12, pp. 41–45, 1997.

[48] E. Schwimmer, "Exchange in the Social Structure of the Orokaiva: Traditional and Emergent Ideologies in the Northern District of Papua", C Hurst and Co-Publishers Ltd., 1973.

[49] M. Pearson and L. Michell, "Smoke Rings: Social Network Analysis of Friendship Groups, Smoking and Drug-taking", *Drugs: Education, Prevention and Policy*, vol. 7, no. 1, pp. 21–37, 2000. http://dx.doi.org/10.1080/dep.7.1.21.37

[50] T. Nepusz *et al.*, "Fuzzy Communities and the Concept of Bridgeness in Complex Networks", *Physical Review E*, vol. 77, no. 1, 016107, 2008. http://dx.doi.org/10.1103/PhysRevE.77.016107

[51] L. C. Freeman *et al.*, "Exploring Social Structure using Dynamic Three-Dimensional Color Images", *Social Networks*, vol. 20, no. 2, pp. 109–118, 1998. http://dx.doi.org/10.1016/S0378-8733(97)00016-6

[52] L. C. Freeman *et al.*, "How Humans See Social Groups: A Test of the Sailer-Gaulin Models", *Journal of Quantitative Anthropology*, vol. 1, pp. 229–238, 1989.

[53] D. A. Smith and D. R. White, "Structure and Dynamics of the Global Economy: Network Analysis of International Trade 1965 – 1980", *Social Forces*, vol. 70, no. 4, pp. 857–893, 1992.

[54] N. Meghanathan and R. Lawrence, "Centrality Analysis of the United States Network Graph", *Proceedings of the 3rd International Conference on Electrical, Electronics, Engineering Trends, Communication, Optimization and Sciences*, Tadepalligudem, India, 2016, pp. 23–28.

[55] H. Jeong *et al.*, "Lethality and Centrality in Protein Networks", *Nature*, vol. 411, pp. 41–42, 2001. http://dx.doi.org/10.1038/35075138

[56] M. Aprahamian *et al.*, "Matching Exponential-based and Resolvent-based Centrality Measures", *Journal of Complex Networks*, vol. 4, no. 2, pp. 157–176, 2016. http://dx.doi.org/10.1093/comnet/cnv016

[57] M. Sajitz-Hermstein and Z. Nikoloski, "Functional Centrality as a Predictor of Shifts in Metabolic Flux States", *BMC Research Notes*, vol. 9, no. 317, pp. 1–4, 2016. http://dx.doi.org/10.1186/s13104-016-2117-0

[58] M. Sajitz-Hermstein and Z. Nikoloski, "Restricted Cooperative Games on Metabolic Networks Reveal Functionally Important Reactions", *Journal of Theoretical Biology*, vol. 314, pp. 192–203, 2012. http://dx.doi.org/10.1016/j.jtbi.2012.08.018

[59] R. Wang *et al.*, "Network-Based Analysis of Software Change Propagation", *The Scientific World Journal,* vol. 2014, article id 237243, pp. 1–10, 2014. http://dx.doi.org/10.1155/2014/237243

[60] D. Koschutzki and F. Schreiber, "Centrality Analysis Methods for Biological Networks and their Application to Gene Regulatory Networks", *Gene Regulation and Systems Biology*, vol. 2008, no. 2, pp. 193–201, 2008.

[61] T. Dawyer *et al.*, "Visual Analytics of Network Centralities", *Proceedings of the Asia-Pacific Symposium on Information Visualization*, vol. 60, pp. 189–197, Tokyo, Japan, 2006.

[62] G. Nomikos *et al.*, "Comparative Assessment of Centrality Indices and Implications on the Vulnerability of ISP Networks", *Proceedings of the 26th International Teletraffic Congress*, Karlskrona, Sweden, 2014, pp. 1–9. http://dx.doi.org/10.1109/ITC.2014.6932932

[63] M. Benzi *et al.*, "Ranking Hubs and Authorities using Matrix Functions", *Linear Algebra and its Applications*, vol. 438, pp. 2447–2474, 2013. https://doi.org/10.1016/j.laa.2012.10.022

[64] M. Grossman *et al.*, "A Survey of Sparse Matrix-Vector Multiplication Performance on Large Matrices", arXiv: 1608.00636.

Received: December 2016
Revised: May 2017
Accepted: May 2017

Contact address:
Natarajan Meghanathan
Department of Computer Science
College of Science, Engineering and Technology
Jackson State University
Jackson, MS, USA
e-mail: natarajan.meghanathan@jsums.com

NATARAJAN MEGHANATHAN received his B. Tech degree in Chemical Engineering from Anna University, India, the MSc degree in Computer Science from Auburn University, AL, USA and PhD degree in Computer Science from The University of Texas in Dallas, USA in 1998, 2002 and 2005 respectively. He joined Jackson State University in 2005 as an Assistant Professor and he is currently a tenured Full Professor in the Department of Computer Science. His research interests are network science, wireless ad hoc and sensor networks, machine learning and cyber security.