

# A Binary Search Algorithm for Correlation Study of Decay Centrality vs. Degree Centrality and Closeness Centrality

Natarajan Meghanathan<sup>1</sup>

<sup>1</sup> Department of Computer Science, Jackson State University, USA

Correspondence: Natarajan Meghanathan, Department of Computer Science, Mailbox 18839, Jackson State University, Jackson, MS 39217, USA. Tel: 1-601-979-3661. E-mail: natarajan.meghanathan@jsums.edu

Received: March 27, 2017

Accepted: April 18, 2017

Online Published: April 25, 2017

doi:10.5539/cis.v10n2p52

URL: <http://doi.org/10.5539/cis.v10n2p52>

*The research is financed by the NASA EPSCoR sub award (#: NNX14AN38A) from University of Mississippi.*

## Abstract

Results of correlation study (using Pearson's correlation coefficient, PCC) between decay centrality (DEC) vs. degree centrality (DEG) and closeness centrality (CLC) for a suite of 48 real-world networks indicate an interesting trend: PCC(DEC, DEG) decreases with increase in the decay parameter  $\delta$  ( $0 < \delta < 1$ ) and PCC(DEC, CLC) decreases with decrease in  $\delta$ . We make use of this trend of monotonic decrease in the PCC values (from both sides of the  $\delta$ -search space) and propose a binary search algorithm that (given a threshold value  $r$  for the PCC) could be used to identify a value of  $\delta$  (if one exists, we say there exists a positive  $\delta$ -space<sub>r</sub>) for a real-world network such that  $\text{PCC}(\text{DEC}, \text{DEG}) \geq r$  as well as  $\text{PCC}(\text{DEC}, \text{CLC}) \geq r$ . We show the use of the binary search algorithm to find the maximum Threshold PCC value  $r_{max}$  (such that  $\delta$ -space <sub>$r_{max}$</sub>  is positive) for a real-world network. We observe a very strong correlation between  $r_{max}$  and PCC(DEC, CLC) as well as observe real-world networks with a larger variation in node degree to more likely have a lower  $r_{max}$  value and vice-versa.

**Keywords:** Correlation Coefficient, Centrality Metrics, Decay Centrality, Binary Search Algorithm, Decay Parameter

## 1. Introduction

The Decay Centrality (DEC) metric is a parameter-driven centrality metric that has not been explored much in the literature for complex network analysis. Decay centrality is a measure of the closeness of a node to the rest of the nodes in the network (Jackson, 2010). However, unlike closeness centrality (CLC; Freeman, 1979), the importance given to the distance (typically, in terms of the number of hops if the edges do not have weights) is weighted in terms of a parameter called the decay parameter  $\delta$  ( $0 < \delta < 1$ ). The formulation for computing the decay centrality of a vertex  $v_i$  for a particular value of the decay parameter  $\delta$  is (Jackson, 2010); see Section 2.3. The decay parameter  $\delta$  essentially controls how important is a node  $v_j$  to a node  $v_i$  ( $v_i \neq v_j$ ) that are at a distance  $d(v_i, v_j)$  from each other. If  $\delta$  is smaller, the distance to the nearby nodes is weighted significantly larger than the distance to the nodes farther away. If  $\delta$  is larger, the distance to every node is given almost the same importance. As a result, if  $\delta$  is closer to 0, the decay centrality of the vertices is more likely to exhibit a very strong positive correlation with the degree centrality of the vertices; if  $\delta$  is closer to 1, the decay centrality of the vertices is more likely to exhibit a very strong positive correlation with the closeness centrality of the vertices.

The motivation for our research came from the initial results (see Figures 1 and 16 for sample results) of our correlation study (conducted with a precision level of  $\epsilon = 0.01$ ) which indicated that the Pearson's correlation coefficient  $\text{PCC}(\text{DEC}_\delta, \text{DEG})$  decreases with increase in  $\delta$  from 0.01 to 0.99 and  $\text{PCC}(\text{DEC}_\delta, \text{CLC})$  decreases with decrease in  $\delta$  from 0.99 to 0.01. Such a trend was observed for all the 48 real-world networks with spectral radius ratio for node degree (Meghanathan, 2014) ranging from 1.01 to 5.51 used in the correlation study. In this paper, we show that this trend could be exploited by developing an efficient binary search algorithm to determine (given a threshold PCC value of  $r$ ) whether there exists a  $\delta$  value for which  $\text{PCC}(\text{DEC}_\delta, \text{DEG})$  as well as  $\text{PCC}(\text{DEC}_\delta, \text{CLC})$  are both greater than or equal to  $r$ . If such a  $\delta$  value is found to exist, we say that there is a positive  $\delta$ -space<sub>r</sub> for the real-world network with respect to the threshold PCC ( $r$ ) for DEC vs. DEG and CLC. We demonstrate the use of the binary search algorithm to determine the maximum value for the threshold PCC

( $r_{max}$ ) for a real-world network such that  $\delta\text{-space}_{max}$  is positive. Our approach is significantly efficient compared to the brute-force approach of computing the DEC values for all possible values of  $\delta$  (note that the  $\delta$ -search space is a continuous space rather than discrete).

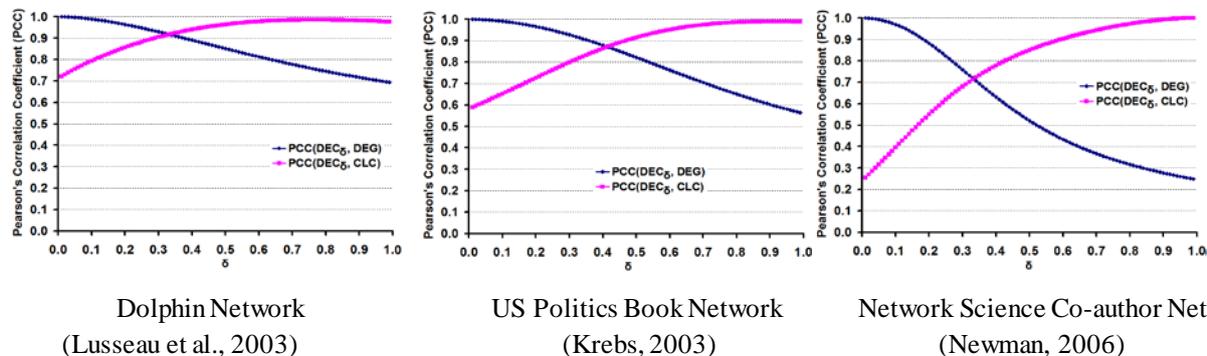


Figure 1. Sample  $PCC(DEC_{\delta}, DEG)$  vs.  $PCC(DEC_{\delta}, CLC)$  Distributions of Real-World Networks

The rest of the paper is organized as follows: In Section 2, we review the centrality metrics (DEG, CLC and DEC) and the Pearson's correlation measure as well as explain their computation with an example graph. Section 3 first introduces the notion of  $\delta\text{-space}_r$  for a threshold PCC ( $r$ ) for DEC-DEG and DEC-CLC correlation (and its computation on the example graph of Section 2). Section 4 describes the proposed binary search algorithm to search for a  $\delta$ -value in  $\delta\text{-space}_r$ , and illustrates its execution with the running example graph of Sections 2-3 for a successful search and an unsuccessful search. Section 4 also illustrates the use of the proposed binary search algorithm to determine the maximum threshold PCC ( $r_{max}$ ) for a network. Section 5 introduces the real-world networks that are analyzed in this paper. Section 6 presents the  $\delta\text{-space}_r$  values and the  $r_{max}$  values for the 48 real-world networks obtained as a result of executing the binary search algorithm. Section 6 also compares the performance of the binary search algorithm vis-a-vis the brute force approach with respect to the number of decay centrality computations needed before deciding whether a real-world network has a positive  $\delta\text{-space}_r$ , or not. Section 7 discusses related work and highlights the contributions of our paper. Section 8 concludes the paper.

**2. Review of Centrality Metrics and Pearson's Correlation Measure**

The centrality metrics that are of interest in this research are degree centrality (DEG), closeness centrality (CLC) and decay centrality (DEC). In this section, we briefly review these three metrics and their computation using a running example graph as well as review the Pearson's correlation measure and its computation with respect to the DEG and CLC metrics for the running example graph.

**2.1 Degree Centrality**

The degree centrality (DEG) of a vertex is the number of neighbors incident on the vertex. Figure 2 illustrates the degree centrality of the vertices (listed above the vertices) in the example graph used in Sections 2-3. 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 2, we observe five of the nine vertices to have a degree of 3). It takes  $\Theta(V)$  time to go through the adjacency list of each vertex; hence, it takes  $\Theta(V^2)$  time to compute the degree centrality for a graph of  $V$  vertices.

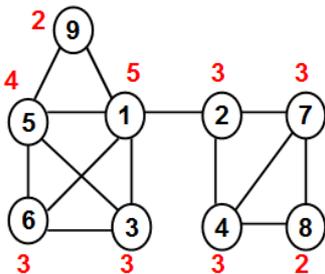


Figure 2. Degree Centrality of the Vertices in an Example Graph

### 2.2 Closeness Centrality

The closeness centrality (CLC) of a vertex (Freeman, 1979) 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  $\Theta(V+E)$ -Breadth First Search (BFS) algorithm (Cormen et. al., 2009) 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 larger the centrality value for a vertex, more important is the vertex. 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 3 illustrates the distance matrix (hop counts of the shortest paths between any two vertices) for the example graph of Figure 2 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  $1/12 = 0.083$ .

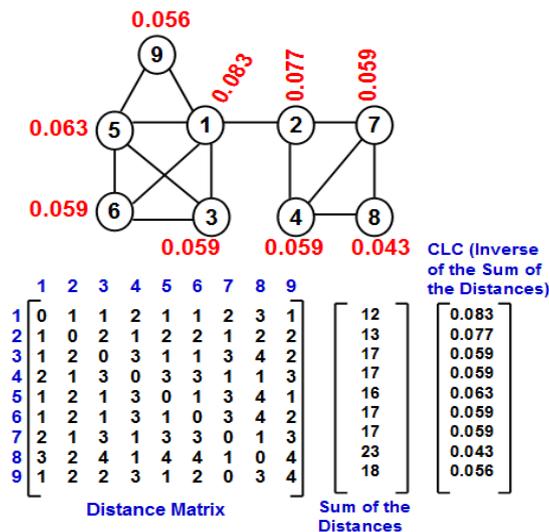


Figure 3. Closeness Centrality of the Vertices in an Example Graph

### 2.3 Decay Centrality

Decay centrality (DEC) is a measure of the closeness of a node to the rest of the nodes in the network (Jackson, 2010). However, unlike closeness centrality, the importance given to the distance (typically, in terms of the number of hops if the edges do not have weights) is weighted in terms of a parameter called the decay parameter  $\delta$  ( $0 < \delta < 1$ ). The formulation for computing the decay centrality of a vertex  $v_i$  for a particular value of the decay parameter  $\delta$  is (Jackson, 2010):

$$DEC(v_i) = \sum_{v_i \neq v_j} \delta^{d(v_i, v_j)} \text{ where } d(v_i, v_j) \text{ is the distance from node } v_i \text{ to node } v_j.$$

The decay parameter  $\delta$  essentially controls how important is a node  $v_j$  to a node  $v_i$  ( $v_i \neq v_j$ ) that are at a distance  $d(v_i, v_j)$  from each other. Nodes that have a higher decay centrality are more likely to be nodes that have several neighbors as well as be much closer to nodes to the rest of the nodes in the network (Tsakas, 2016). Figure 4 presents the decay centrality of the vertices in the example graph of Section 2 for different values of the decay parameter  $\delta$ . We also illustrate sample calculations of the decay centrality of vertex 1 for three different values of  $\delta$ .

On a graph of  $V$  vertices and  $E$  edges, it takes  $\Theta(V^3)$  time to compute the distance matrix (the shortest path weights between any two nodes in the network) using the Floyd-Warshall algorithm (Cormen et. al., 2009) for weighted graphs and  $\Theta(V(V+E))$  time to compute the distance matrix (the hop count of the shortest paths between any two nodes in the network) using the  $\Theta(V+E)$ -BFS algorithm for graphs with unit edge weights. Since  $E = \Theta(V^2)$ , we could say, in general, it takes  $\Theta(V^3)$  time to compute the distance matrix for any graph. Given the distance matrix for a graph, it takes  $\Theta(V)$  time to compute the decay centrality of a particular vertex  $v_i$

as we have to find  $\delta^{d(v_i, v_j)}$  for every vertex  $v_j \neq v_i$ . Hence, given the distance matrix for a graph, it would take  $\Theta(V^2)$  time to compute the decay centrality of the vertices. Overall, given a graph of  $V$  vertices, the time-complexity to compute decay centrality is  $\Theta(V^3) + \Theta(V^2) = \Theta(V^3)$ . Thus, the time-complexity to compute the decay centrality of the vertices is dominated by the time-complexity to compute the distance matrix.

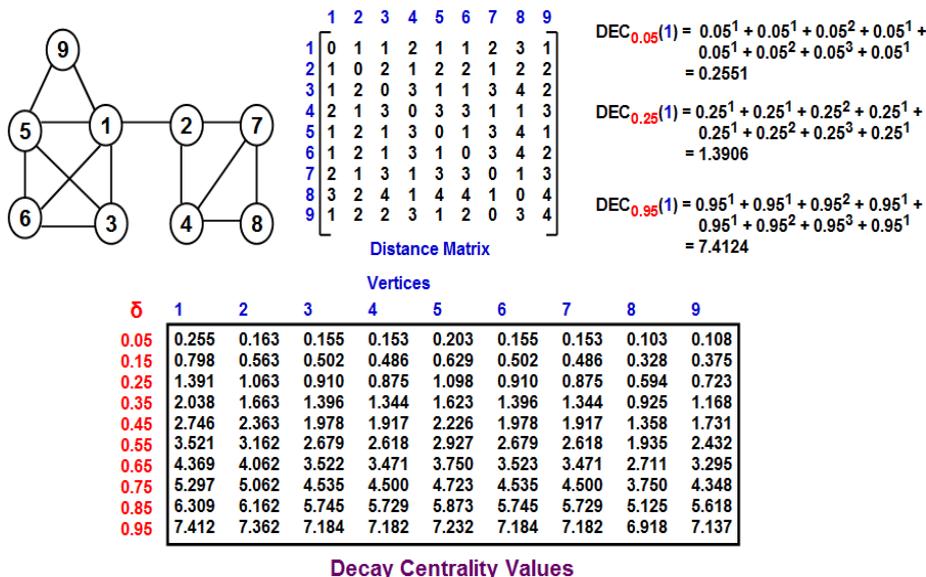


Figure 4. Decay Centrality of the Vertices in an Example Graph

2.4 Pearson's Correlation Measure

We use the Pearson's correlation coefficient (PCC; Lay et al., 2015) as the measure for analyzing the correlation between the decay centrality (computed for different values of the decay parameter  $\delta$ ) and the degree centrality and closeness centrality. The Pearson's product-moment correlation when applied for centrality metrics is a measure of the linear dependence between any two metrics in consideration (Lay et al., 2015). 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 below to calculate the correlation coefficient. If  $X$  and  $Y$  are the datasets for two centrality metrics: let  $X_i$  and  $Y_i$  indicate the centrality values for the individual vertices  $v_i$  ( $1 \leq i \leq n$ , where  $n$  is the number of vertices) and  $\bar{X}$  and  $\bar{Y}$  are the average of the centrality values;  $PCC(X, Y)$  is calculated as follows. Figure 5 illustrates the computation of the Pearson's correlation coefficient between DEG and CLC.

$$PCC(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \tag{1}$$

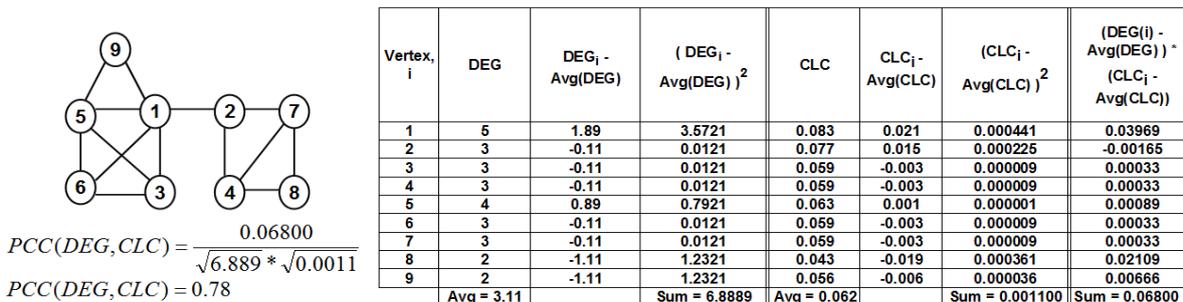


Figure 5. Sample Illustration of the Computation of the Pearson's Correlation Coefficient between Degree Centrality and Closeness Centrality

### 3. $\delta$ -Space for DEG-DEC-CLC Correlation

Figure 6 displays the Pearson's correlation coefficient values between each of the two centrality metrics (DEG and CLC) and the decay centrality values ( $DEC_\delta$ ) computed for different values of the decay parameter  $\delta$  ranging from 0.01 to 0.99 for the example graph of Figures 2-5. We see  $PCC(DEC_\delta, DEG)$  to monotonically decrease with increase in  $\delta$  and  $PCC(DEC_\delta, CLC)$  to monotonically decrease with decrease in  $\delta$ . A similar trend is also noticed for all the 48 real-world network graphs analyzed in Section 6. Using this as the basis, we define the  $\delta$ -space<sub>r</sub> for a real-world network with respect to a threshold PCC ( $r$ ) as the difference between the maximum  $\delta$  value for which we observe  $PCC(DEC_\delta, DEG) \geq r$  and the minimum  $\delta$  value for which we observe  $PCC(DEC_\delta, CLC) \geq r$ .

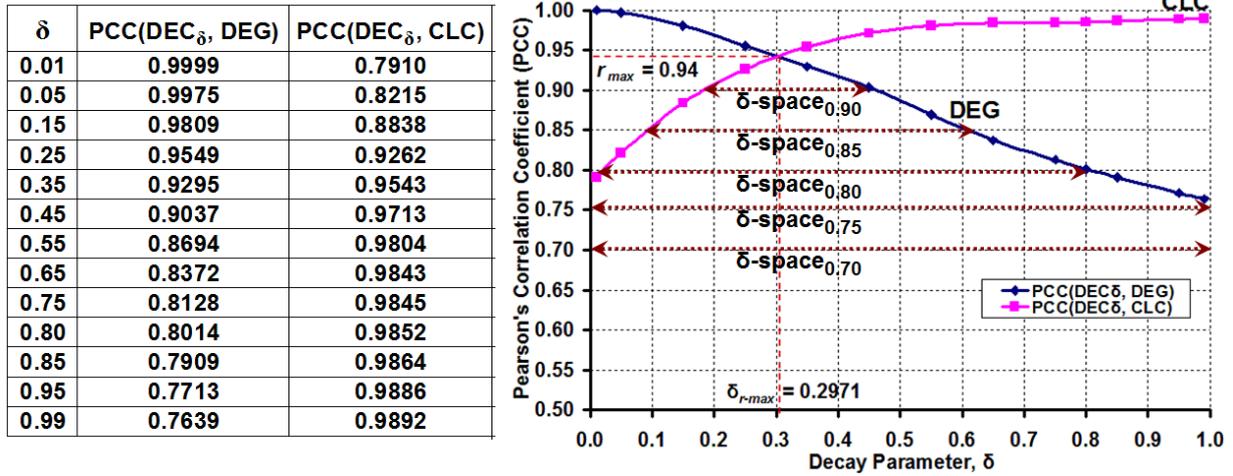


Figure 6. Distribution of the Pearson's Correlation Coefficient Values between Decay Centrality and the Two Centrality Metrics (Degree Centrality and Closeness Centrality) vs. the Decay Parameter  $\delta$  for the Example Graph of Figures 2-5

$\delta$ -space<sub>r</sub> for a threshold PCC ( $r$ ) basically quantifies the range of values in the open interval  $[\delta_{PCC(DEC,CLC) \geq r}^{\min} \dots \delta_{PCC(DEC,DEG) \geq r}^{\max}]$  that could be chosen from the closed interval (0...1) to determine decay centrality values that exhibit PCC of the threshold value of  $r$  or above with both degree and closeness centralities. Note that we did not choose the interval (0, 1] for  $\delta$ , as  $\delta = 1$  would correspond to the component size and not quantify the centrality of the vertices. On the same lines, we did not choose the interval [0, 1) for  $\delta$ , as  $\delta = 0$  would make the decay centrality of the vertices to become zero. Quantitatively,  $\delta$ -space<sub>r</sub> is defined as follows, where  $\epsilon$  corresponds to the level of precision used for  $\delta$  in the range (0...1). Note that  $\delta$ -space<sub>r</sub> could be determined for any threshold value of the Pearson's correlation coefficient ( $r$ ) of interest.

$$\delta\text{-space}_r = \delta_{PCC(DEC,DEG) \geq r}^{\max} - \delta_{PCC(DEC,CLC) \geq r}^{\min} + \epsilon \quad \text{if } \delta_{PCC(DEC,CLC) \geq r}^{\min} \leq \delta_{PCC(DEC,DEG) \geq r}^{\max} \quad (2)$$

$$\delta\text{-space}_r = \delta_{PCC(DEC,DEG) \geq r}^{\max} - \delta_{PCC(DEC,CLC) \geq r}^{\min} \quad \text{if } \delta_{PCC(DEC,CLC) \geq r}^{\min} > \delta_{PCC(DEC,DEG) \geq r}^{\max} \quad (3)$$

Note that if  $\delta_{PCC(DEC,CLC) \geq r}^{\min} = \delta_{PCC(DEC,DEG) \geq r}^{\max}$ , it implies there is one  $\delta$  value ( $\delta = \delta_{\min,r}^{DEC-CLC} = \delta_{\max,r}^{DEC-DEG}$ ) for which  $PCC(DEC_\delta, DEG) \geq r$  and  $PCC(DEC_\delta, CLC) \geq r$ . If  $\delta_{PCC(DEC,CLC) \geq r}^{\min} < \delta_{PCC(DEC,DEG) \geq r}^{\max}$ , there might be more than one  $\delta$  value that could be chosen from the open interval  $[\delta_{PCC(DEC,CLC) \geq r}^{\min} \dots \delta_{PCC(DEC,DEG) \geq r}^{\max}]$ ; hence, we add the precision level  $\epsilon$  in the formulation for  $\delta$ -space<sub>r</sub> when  $\delta_{PCC(DEC,CLC) \geq r}^{\min} \leq \delta_{PCC(DEC,DEG) \geq r}^{\max}$ . On the other hand, if  $\delta_{PCC(DEC,CLC) \geq r}^{\min} > \delta_{PCC(DEC,DEG) \geq r}^{\max}$ , it implies there is not even one single  $\delta$  value for which  $DEC_\delta$  would exhibit the threshold PCC of  $r$  or above with both DEG and CLC. Hence, we do not add the precision level  $\epsilon$  in the  $\delta$ -space<sub>r</sub> formulation for  $\delta_{PCC(DEC,CLC) \geq r}^{\min} > \delta_{PCC(DEC,DEG) \geq r}^{\max}$ .

Among the  $\delta$  values (precision  $\epsilon = 0.01$ ) shown in Figure 6, the largest  $\delta$  value for which  $PCC(DEC_\delta, DEG) \geq 0.80$  is  $\delta_{PCC(DEC,DEG) \geq r=0.80}^{\max} = 0.80$  and the smallest  $\delta$  value for which  $PCC(DEC_\delta, CLC) \geq 0.80$  is

$\delta_{PCC(DEC,CLC)\geq r=0.80}^{\min} = 0.03$ . Hence, the  $\delta\text{-space}_{r=0.8}$  for the example graph is  $\delta_{PCC(DEC,DEG)\geq r=0.80}^{\max} - \delta_{PCC(DEC,DEG)\geq r=0.80}^{\max} + \epsilon = 0.80 - 0.03 + 0.01 = 0.78$ . Note that 0.78 is also  $PCC(DEC, CLC)$  for the example graph (computed in Figure 5).

Table 1 displays the  $\delta\text{-space}_r$  values for different values of the threshold  $PCC(r)$ , ranging from 0.70 to 0.95 for the running example graph of Figures 2-5. Note that  $\delta\text{-space}_r$  decreases as the threshold  $PCC(r)$  value increases. We say  $\delta\text{-space}_r$  does not exist for a network graph if it is less than zero. For the running example graph, we observe  $\delta\text{-space}_r$  to exist (as shown in Figures 6 and 7) for  $r = 0.70 \dots 0.90$  (in increments of 0.05); but,  $\delta\text{-space}_{r=0.95}$  does not exist. Figures 6 and 7 illustrate the shrinking  $\delta\text{-space}_r$  for the example network graph with increase in the threshold  $PCC(r)$  value. We observe the DEG centrality to be relatively a bottleneck metric (vis-a-vis CLC) that primarily contributes to the shrinkage of  $\delta\text{-space}_r$ . A similar trend is also observed for real-world networks whose  $\delta\text{-space}_r$  is less than the maximum possible value of  $1 - \epsilon$ . Figure 7 also clearly illustrates that  $\delta\text{-space}_r$  is the overlap of the intervals  $[0 + \epsilon \dots \delta_{PCC(DEC,DEG)\geq r}^{\max}]$  and  $[\delta_{PCC(DEC,CLC)\geq r}^{\min} \dots 1 - \epsilon]$ . We make use of this observation as part of the design logic for the binary search algorithm (described in Section 4) to determine the existence of  $\delta\text{-space}_r$ .

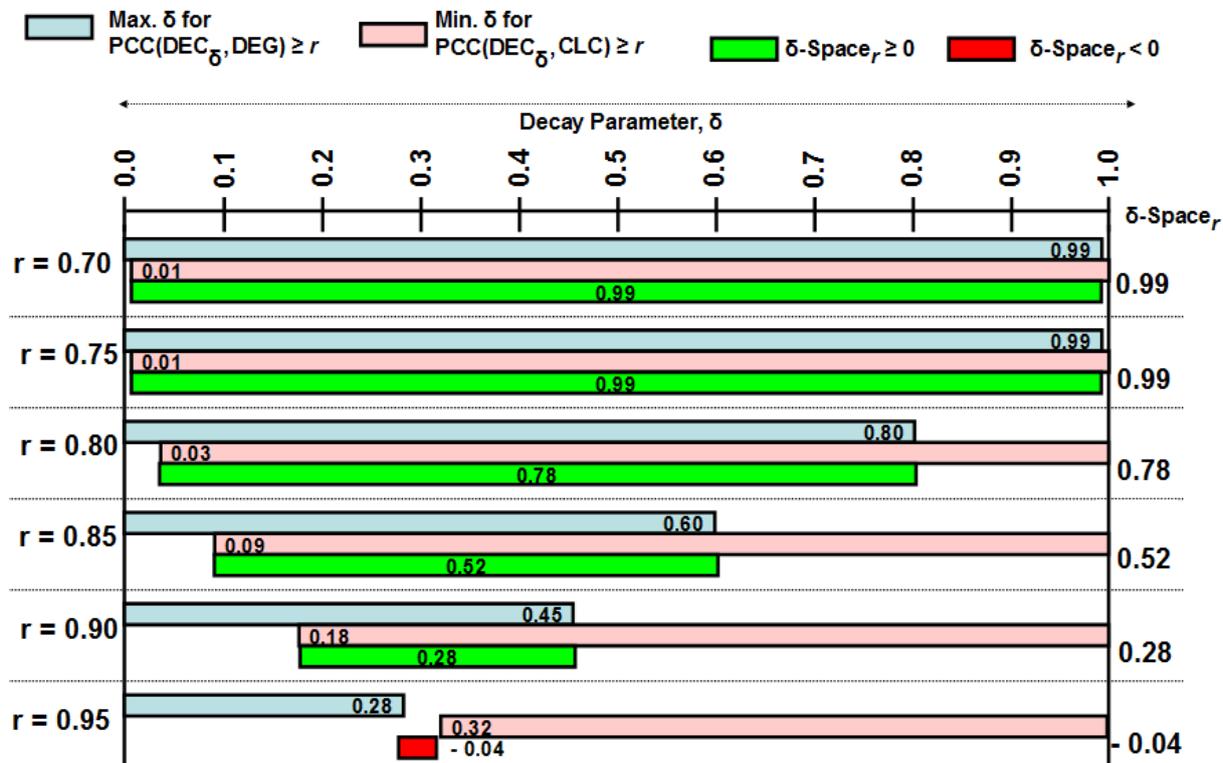


Figure 7. Shrinking  $\delta\text{-Space}_r$  with Increase in Threshold  $PCC(r)$  for the Example Graph of Figures 2-5

Table 1.  $\delta\text{-space}_r$  Values for the Example Graph of Figures 2-5 for Different Values of Threshold  $PCC(r)$

Threshold $PCC(r)$	$\delta_{PCC(DEC,DEG)\geq r}^{\max}$	$\delta_{PCC(DEC,CLC)\geq r}^{\min}$	$\delta\text{-space}_r$	$\delta\text{-space}_r$ Exists?
0.70	0.99	0.01	0.99	YES
0.75	0.99	0.01	0.99	YES
0.80	0.80	0.03	0.78	YES
0.85	0.60	0.09	0.52	YES
0.90	0.45	0.18	0.28	YES
0.95	0.28	0.32	-0.04	NO

#### 4. Binary Search Algorithm to Determine the Existence of $\delta$ -Space<sub>r</sub> and Maximum Threshold PCC for DEC-DEG and DEC-CLC Correlation

We now describe a binary search algorithm (see Figure 8 for the pseudo code) whose objective is (given a threshold PCC value of  $r$ ) to find a particular value of the decay parameter  $\delta$  (if one exists) such that  $PCC(DEC_{\delta}, DEG) \geq r$  and  $PCC(DEC_{\delta}, CLC) \geq r$ . If a  $\delta$  value could be found for a real-world network, it implies there does exist a overlap of the intervals  $[0 + \varepsilon \dots \delta_{PCC(DEC, DEG) \geq r}^{\max}]$  and  $[\delta_{PCC(DEC, CLC) \geq r}^{\min} \dots 1 - \varepsilon]$  and we could conclude that the  $\delta$ -space<sub>r</sub> for the real-world network exists (i.e.,  $\delta$ -space<sub>r</sub> > 0). If a  $\delta$  value could not be found, it implies the intervals  $[0 + \varepsilon \dots \delta_{PCC(DEC, DEG) \geq r}^{\max}]$  and  $[\delta_{PCC(DEC, CLC) \geq r}^{\min} \dots 1 - \varepsilon]$  do not overlap ( $\delta_{PCC(DEC, CLC) \geq r}^{\min} > \delta_{PCC(DEC, DEG) \geq r}^{\max}$ ) and the  $\delta$ -space<sub>r</sub> for the real-world network does not exist.

**Given:** Threshold PCC  $r$ , Precision Level  $\in$

**Auxiliary Variables:** Left Index, Right Index, Middle Index,  $\delta$

**Initialization:** Left Index =  $\in$ , Right Index = 1

**Output:**  $\delta$

**Begin Binary Search Algorithm**

**If** (  $PCC(DEC_{\delta = \text{Left Index}}, CLC) \geq r$  ) **then**

**return**  $\delta = \text{Left Index}$

**End If**

**While** (Left Index  $\leq$  Right Index) **do**

$Middle\ Index = \lfloor (LeftIndex + RightIndex) / 2 \rfloor$

**If** (  $PCC(DEC_{\delta = Middle\ Index}, CLC) \geq r$  AND  $PCC(DEC_{\delta = Middle\ Index}, DEG) \geq r$  ) **then**

**return**  $\delta = Middle\ Index$

**End If**

**If** (  $PCC(DEC_{\delta = Middle\ Index}, CLC) \geq r$  AND  $PCC(DEC_{\delta = Middle\ Index}, DEG) < r$  ) **then**

$Left\ Index = Middle\ Index$

**End If**

**If** (  $PCC(DEC_{\delta = Middle\ Index}, CLC) < r$  AND  $PCC(DEC_{\delta = Middle\ Index}, DEG) \geq r$  ) **then**

$Right\ Index = Middle\ Index$

**End If**

**If** (  $PCC(DEC_{\delta = Middle\ Index}, CLC) < r$  AND  $PCC(DEC_{\delta = Middle\ Index}, DEG) < r$  ) **then**

**return**  $\delta = -1$  // does not exist

**End If**

**End While**

**End Binary Search Algorithm**

Figure 8. Pseudo Code for the Proposed Binary Search Algorithm to Determine the Existence of  $\delta$ -Space<sub>r</sub> and Maximum Threshold PCC for DEC-DEG and DEC-CLC Correlation

The binary search algorithm makes use of the observation in this research that  $PCC(DEC_{\delta}, DEG)$  monotonically decreases as  $\delta$  increases from 0 to 1, and  $PCC(DEC_{\delta}, CLC)$  monotonically decreases as  $\delta$  decreases from 1 to 0. Like the standard binary search algorithm, the binary search algorithm described here also maintains two indices: a left index and a right index, and the two indices approach towards each other during the course of the algorithm. Also, as in the standard binary search algorithm, we exit from the iterations (described below) when the right

index becomes larger than the left index. An invariant in the binary search algorithm is that the left index is a  $\delta$  value for which  $PCC(DEC_{\delta}, DEG) \geq \text{threshold } r$  and the right index is a  $\delta$  value for which  $PCC(DEC_{\delta}, CLC) \geq \text{threshold } r$ . As part of optimization, before proceeding to the iterations of the algorithm, one can test whether  $PCC(DEC_{\delta=\text{initial left index}}, CLC) \geq \text{threshold } r$ . If so, we are done and the  $\delta$  value of interest could be the value of the initial left index itself. This would be especially useful for lower values of the threshold PCC, as we observe  $\delta_{PCC(DEC, CLC) \geq r}^{\min}$  to be closer to 0.01 for lower values for the threshold PCC  $r$ .

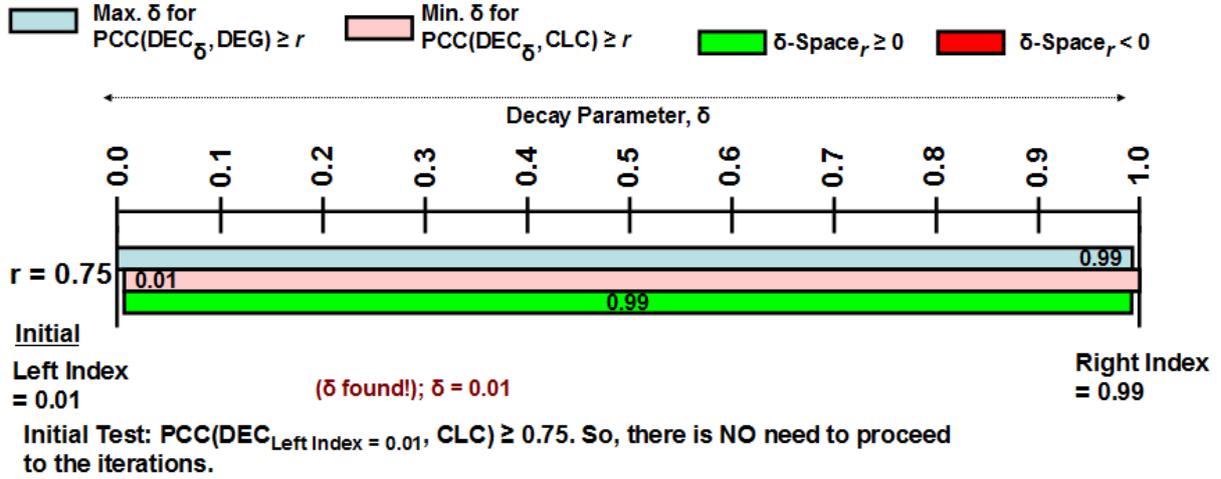


Figure 9. Example to Illustrate the Execution of the Optimization Step (Initial Test with the Left Index) of the Binary Search Algorithm to Determine the Existence of  $\delta\text{-space}_r$  (Threshold PCC  $r = 0.75$ )

In each iteration of the binary search algorithm, we find a middle index that is the average of the left index and right index and test whether  $PCC(DEC_{\delta=\text{middle index}}, DEG) \geq \text{threshold } r$  and  $PCC(DEC_{\delta=\text{middle index}}, CLC) \geq \text{threshold } r$ . If so, we are done and the  $\delta$  value corresponding to the middle index is the  $\delta$  value of interest. Otherwise, we do the following three tests:

- If  $PCC(DEC_{\delta=\text{middle index}}, DEG) \geq r$  and  $PCC(DEC_{\delta=\text{middle index}}, CLC) < r$ , we move the left index to the middle index (as the  $DEC_{\delta}$ -CLC correlation coefficient value is less than  $r$  for  $\delta = \text{middle index}$  and it can only decrease further if the  $\delta$ -search space is moved to the left of the middle index; hence, we shrink the  $\delta$ -search space to the right of the middle index).
- If  $PCC(DEC_{\delta=\text{middle index}}, DEG) < r$  and  $PCC(DEC_{\delta=\text{middle index}}, CLC) \geq r$ , we move the right index to the middle index (as the  $DEC_{\delta}$ -DEG correlation coefficient is less than  $r$  for  $\delta = \text{middle index}$  and it can only decrease further if we move to the right of the middle index; hence, we shrink the  $\delta$ -search space to the left of the middle index).
- If  $PCC(DEC_{\delta=\text{middle index}}, DEG) < r$  and  $PCC(DEC_{\delta=\text{middle index}}, CLC) < r$ , we exit the algorithm and declare that there is no single  $\delta$  value for which  $DEC_{\delta}$  would exhibit the threshold value of correlation coefficient or above with both DEG and CLC. This is because: as per the observation  $DEC_{\delta=\text{middle index}}$  not exhibiting the threshold value of correlation coefficient with DEG, the  $\delta$  value of interest has to be in the range [left index ... middle index); whereas, as per the observation  $DEC_{\delta=\text{middle index}}$  not exhibiting the threshold value of correlation coefficient with CLC, the  $\delta$  value of interest has to be in the range (middle index ... right index]. The two ranges do not overlap and hence there cannot be a  $\delta$  value for which  $DEC_{\delta}$  exhibits the threshold value of correlation coefficient with both DEG and CLC.

The binary search algorithm could be executed over a  $\delta$ -search space of  $[0+\epsilon \dots 1-\epsilon]$ , wherein the initial values for the left index and right index are respectively  $0+\epsilon$  and  $1-\epsilon$ . The value of  $\epsilon$  used in this paper is 0.01. Figure 9 presents a scenario wherein there is no need to proceed to the iterations of the binary search algorithm and a simple initial test is sufficient to determine the existence of  $\delta\text{-space}_r$ . The initial test is to check whether  $PCC(DEC_{\delta=\text{initial left index}} = 0.01, CLC) \geq \text{threshold PCC of } r (= 0.75, \text{ in Figure 9})$ . Since,  $PCC(DEC_{0.01}, CLC) \geq 0.75$  and we already know that  $PCC(DEC_{0.01}, DEG) \geq 0.75$ , we can stop right away and declare that the  $\delta$  of interest is 0.01 and that  $\delta\text{-space}_r$  exists for a threshold PCC  $r$  value of 0.75.

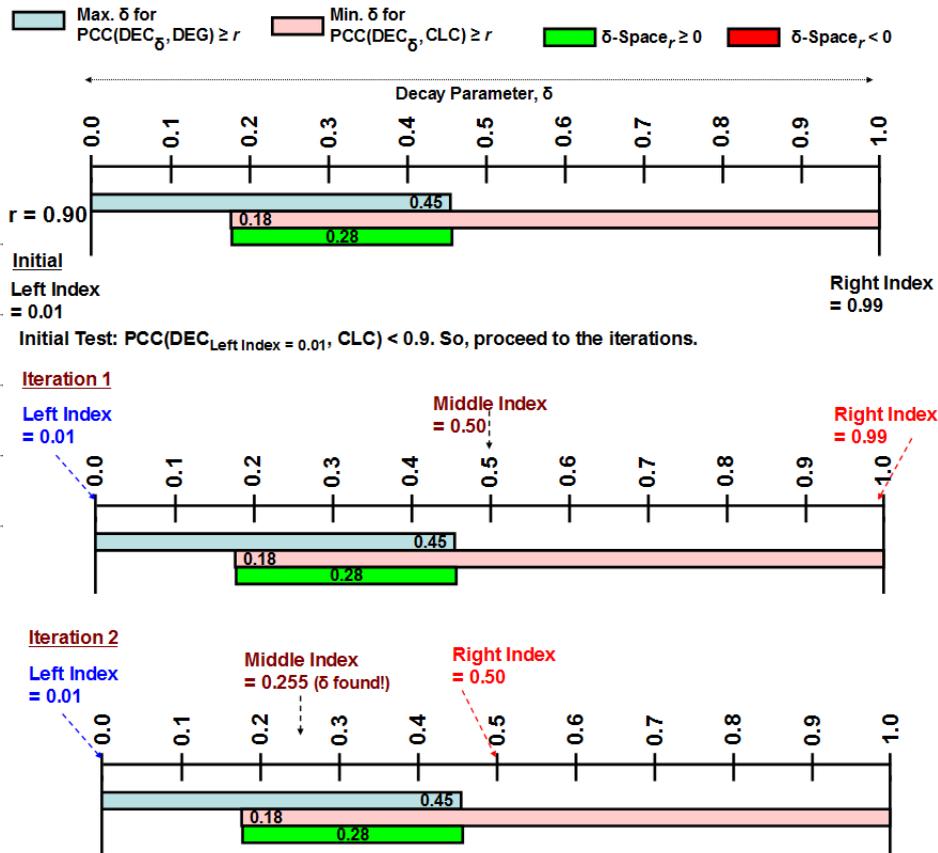


Figure 10. Example to Illustrate the Iterations of a Successful Search of the Binary Search Algorithm to Determine the Existence of  $\delta\text{-space}_r$ , (Threshold PCC  $r = 0.90$ )

Figure 10 presents a scenario for a successful search (threshold PCC value  $r = 0.90$ ) wherein we are required only two iterations of the algorithm to determine the existence of  $\delta\text{-space}_r$ . During the first iteration, we found  $DEC_{\text{middle index} = 0.50}$  not to exhibit a correlation coefficient of 0.90 or above with DEG, but correlated with CLC. Hence, we move the right index to the middle index. During the second iteration, we found  $DEC_{\text{middle index} = 0.25}$  to exhibit a correlation coefficient of 0.90 or above with both DEG and CLC. Thus, the algorithm stops with  $\delta = 0.25$  as the  $\delta$  value of interest and we conclude that  $\delta\text{-space}_{r=0.90}$  exists.

Figure 11 presents a scenario for an unsuccessful search (threshold PCC value  $r = 0.95$ ) wherein we needed four iterations to bring in the middle index to a  $\delta$  value for which both  $PCC(DEC_{\delta}, DEG)$  and  $PCC(DEC_{\delta}, CLC)$  are less than the threshold PCC  $r$  value of 0.95. The [left index ... right index] ranges for the four iterations are respectively: Iteration 1 - [0.01 ... 0.99]; Iteration 2 - [0.01 ... 0.50]; Iteration 3 - [0.255 ... 0.50]; Iteration 4 - [0.255 ... 0.3775]. The middle index at the end of the fourth iteration is  $(0.255 + 0.3775)/2 = 0.316$  and  $DEC_{0.316}$  (i.e., DEC values computed for  $\delta = 0.316$ ) does not exhibit a correlation coefficient of 0.95 or above with both DEG and CLC. Hence, the range for DEC to exhibit a correlation coefficient of 0.95 or above with DEG has to be towards the left of this middle index (i.e.,  $< 0.316$ ) and the range for DEC to exhibit a correlation coefficient of 0.95 or above with CLC has to be towards the right of this middle index (i.e.,  $> 0.316$ ). This is not possible and hence we stop and declare that  $\delta\text{-space}_{r=0.95}$  does not exist for this graph.

The basic operation (the most time consuming step) of the binary search algorithm is the computation of the decay centrality of the vertices in each iteration as well as during the initial test with the left index. Since the  $\delta$ -search space is a continuous search space (i.e., real numbers from 0 to 1), we are not able to theoretically quantify the average and worst-case number of times the basic operation would be executed as part of the binary search algorithm. Nevertheless, as seen in the experimental results presented in Section 6 for real-world networks, we anticipate the number of times the basic operation is executed as part of the binary search algorithm (both the average and worst-case) to be significantly smaller than the number of times the basic operation is executed as part of a brute-force search.

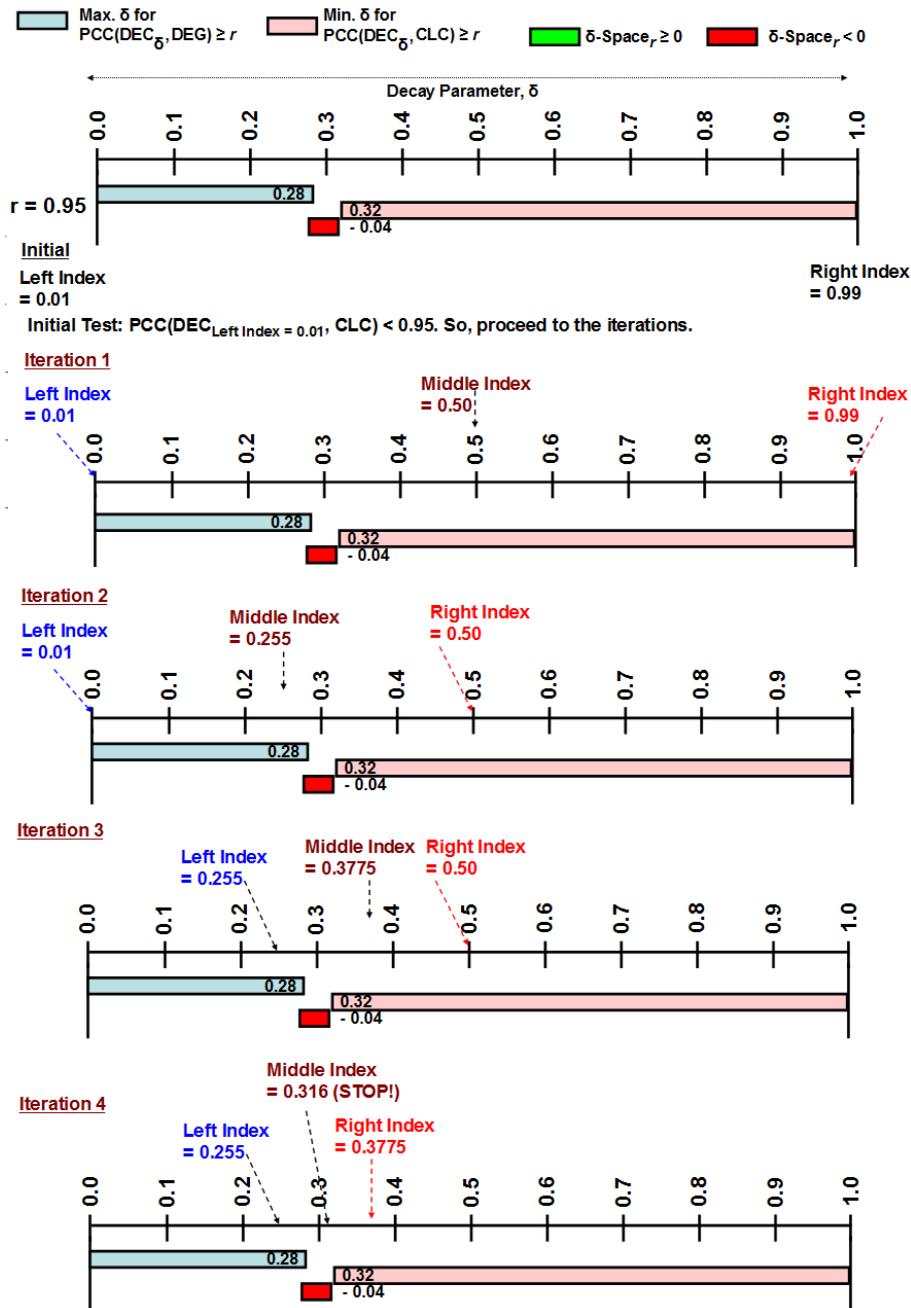


Figure 11. Example to Illustrate the Iterations of an Unsuccessful Search of the Binary Search Algorithm to Determine the Existence of  $\delta\text{-space}_r$  (Threshold PCC  $r = 0.95$ )

We now explain the use of the binary search algorithm to determine the maximum value of the threshold PCC ( $r_{max}$ ) for which there exists a positive  $\delta\text{-space}$ . The idea is to start with a *tentative*  $r_{max}$  value of 1.0 and use the binary search algorithm to test if there exists a positive  $\delta\text{-space}_{tentative\ max}$ . If so, we stop and declare the *tentative*  $r_{max}$  as the *final*  $r_{max}$  value. Otherwise, we reduce the value of *tentative*  $r_{max}$  by 0.01 and continue the above test until we find a *tentative*  $r_{max}$  for which there exists a positive  $\delta\text{-space}$ . We also go ahead and determine the  $\delta$  value (referred to as  $\delta_{max}$ ) that was found to be part of the  $\delta\text{-space}_{max}$ . The  $r_{max}$  value for the example graph in Figures 2-5 is 0.94 and the corresponding  $\delta_{max}$  value is 0.2971. This is evident from Figure 6 in which we show the decrease in the  $PCC(DEC_\delta, DEG)$  values and the increase in  $PCC(DEC_\delta, CLC)$  values with increase in  $\delta$  from 0.01 to 0.99. The values of  $(\delta_{max}, r_{max}) = (0.2971, 0.94)$  correspond to the intersection point between the two PCC curves. The  $\delta_{max}$  value of 0.2971 could be approximated to the value of 0.30 as it appears in Figure 6.

## 5. Real-World Network Graphs

In this section, we introduce the 48 real-world networks analyzed in this paper. Table 2 lists the three character code acronym, name and the network type, the values for the number of nodes and edges as well as the spectral radius ratio for node degree ( $\lambda_{sp}$ ). All the real-world networks are modeled as undirected graphs. The spectral radius ratio for node degree (Meghanathan, 2014) is a measure of the variation in node degree and is calculated as the ratio of the principal eigenvalue (Bonacich, 1987) 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 1.0; the farther is the ratio from the value of 1.0, 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 5.51 (indicating that the real-world network graphs analyzed range from random networks (Renyi, 1959) with smaller variation in node degree to scale-free networks (Barabasi & Albert, 1999) 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): I. Acquaintance network (12), II. Friendship network (9), III. Co-appearance network (6), IV. Employment network (4), V. Citation network (3), VI. Collaboration network (3), VII. Literature network (3), VIII. Political network (2), IX. Biological network (2), X. Game network (2), XI. Transportation network (1) and XII. Trade network (1). A brief description about 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.

Table 2. Real-World Networks used in the Correlation Analysis

#	Net.	Network Name <sup>Type</sup>	Reference	$\lambda_{sp}$	#nodes	#edges
1	ADJ	Word Adjacency Net <sup>III</sup>	Newman, 2006	1.73	112	425
2	AKN	Anna Karenina Net <sup>III</sup>	Knuth, 1993	2.48	140	494
3	JBN	Jazz Band Net <sup>IV</sup>	Geiser & Danon, 2003	1.45	198	2742
4	CEN	C. Elegans Neural Net <sup>IX</sup>	White et. al., 1986	1.68	297	2148
5	CLN	Centrality Literature Net <sup>V</sup>	Hummon et. al., 1990	2.03	118	613
6	CGD	Citation Graph Draw. Net <sup>V</sup>	Biedl & Franz, 2001	2.24	259	640
7	CFN	Copperfield Net <sup>III</sup>	Knuth, 1993	1.83	89	407
8	DON	Dolphin Net <sup>I</sup>	Lusseau et. al., 2003	1.40	62	159
9	DRN	Drug Net <sup>I</sup>	Lee, 2004	2.76	212	284
10	DLN	Dutch Lit. 1976 Net <sup>VII</sup>	Nooy, 1999	1.49	37	81
11	ERD	Erdos Collaboration Net <sup>VI</sup>	Batagelj & Mrvar, 2006	3.00	433	1314
12	FMH	Faux Mesa High Sch. Net <sup>II</sup>	Resnick et. al., 1997	2.81	147	202
13	FHT	Friendship Hi-Tech Firm <sup>II</sup>	Krackhardt, 1999	1.57	33	91
14	FTC	Flying Teams Cade Net <sup>IV</sup>	Moreno, 1960	1.21	48	170
15	FON	US Football Net <sup>X</sup>	Girvan & Newman, 2002	1.01	115	613
16	CDF	College Dorm Fraternity <sup>I</sup>	Bernard et. al., 1980	1.11	58	967
17	GD96	Graph Drawing 1996 Net <sup>V</sup>	Batagelj & Mrvar, 2006	2.38	180	228
18	MUN	Marvel Universe Net <sup>III</sup>	Gleiser, 2007	2.54	167	301
19	GLN	Graph Glossary Net <sup>VII</sup>	Batagelj & Mrvar, 2006	2.01	67	118
20	HTN	Hypertext 2009 Net <sup>I</sup>	Isella et. al., 2011	1.21	115	2164
21	HCN	Huckleberry Co. Net <sup>III</sup>	Knuth, 1993	1.66	76	302
22	ISP	Infec. Socio-Patt. Net <sup>I</sup>	Isella et. al., 2011	1.69	309	1924
22	KCN	Karate Club Net <sup>I</sup>	Zachary, 1977	1.47	34	78
24	KFP	Korea Family Plan. Net <sup>I</sup>	Rogers & Kincaid, 1980	1.70	37	85
25	LMN	Les Miserables Net <sup>III</sup>	Knuth, 1993	1.82	77	254
26	MDN	Macaque Dom. Net <sup>IX</sup>	Takahata, 1991	1.04	62	1167
27	MTB	Madrid Train Bomb. Net <sup>I</sup>	Hayes, 2006	1.95	64	295
28	MCE	Manufact. Comp. Empl <sup>IV</sup>	Cross et. al., 2004	1.12	77	1549
29	MSJ	Soc. Net. Journal Net <sup>VI</sup>	McCarty & Freeman, 2008	3.48	475	625
30	AFB	Author Facebook Net <sup>II</sup>	-	2.29	171	940
31	MPN	Mexican Pol. Elite Net <sup>VIII</sup>	Gil-Mendieta & Schmidt, 1996	1.23	35	117
32	MMN	ModMath Net <sup>II</sup>	Batagelj & Mrvar, 2006	1.59	30	61

33	NSC	Net. Science Co-author <sup>VI</sup>	Newman, 2006	5.51	1,589	2,743
34	PBN	US Politics Books Net <sup>VII</sup>	Krebs, 2003	1.42	105	441
35	PSN	Primary Sch. Contact Net <sup>I</sup>	Gemmetto et. al., 2014	1.22	238	5539
36	PFN	Prison Friendship Net <sup>II</sup>	MacRae, 1960	1.32	67	142
37	SJN	San Juan Sur Family Net <sup>I</sup>	Loomis et al., 1953	1.29	75	155
38	SDI	Scotland Corp. Interlock <sup>IV</sup>	Scott, 1980	1.94	230	359
39	SPR	Senator Press Rel. Net <sup>VIII</sup>	Grimmer, 2010	1.47	92	477
40	SWC	Soccer World Cup Net <sup>X</sup>	Batagelj & Mrvar, 2006	1.45	35	118
41	SSM	Sawmill Strike Comm. Net <sup>I</sup>	Michael et. al., 1997	1.22	24	38
42	TEN	Taro Exchange Net <sup>I</sup>	Schwimmer et al., 1973	1.06	22	39
43	TWF	Teenage Fem. Friend Net <sup>II</sup>	Pearson & Michell, 2000	1.59	47	77
44	UKF	UK Faculty Friend Net <sup>II</sup>	Nepusz et. al., 2008	1.35	83	578
45	APN	US Airports 1997 Net <sup>XII</sup>	Batagelj & Mrvar, 2006	3.22	332	2126
46	RHF	Residence Hall Friend Net <sup>II</sup>	Freeman et. al., 1998	1.27	217	1839
47	WSB	Windsurfers Beach Net <sup>II</sup>	Freeman et. al., 1989	1.22	43	336
48	WTN	World Trade Metal Net <sup>XIII</sup>	Smith & White, 1992	1.38	80	875

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 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 books/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 *trade network* is a network of countries/people involved in certain trade.

## 6. Results of Correlation Study

In this section, we present and analyze the results of our correlation study for the 48 real-world networks. Table 3 lists the  $\delta\text{-space}_r$  values and the number of decay centrality computations (the basic operation) for the binary search algorithm. The values for the threshold PCC in Table 3 range from 0.60 to 0.95. As expected, the  $\delta\text{-space}_r$  values decrease as the value for the threshold PCC ( $r$ ) increases. The median of the  $\delta\text{-space}_r$  values for the different threshold  $r$  values are also given in the bottom of Table 3. From median value 0.99 for  $r = 0.60$ , the median reduces to 0.515 for  $r$  value of 0.80 and to -0.14 for  $r$  value of 0.95. Cells with negative  $\delta\text{-space}_r$  values are highlighted in light pink color. Though the author Facebook network had negative  $\delta\text{-space}_r$  values for  $r$  values starting from 0.60, we notice the negative  $\delta\text{-space}_r$  values for the real-world networks are more prominent for  $r$  values starting from 0.80. More than 40% and 50% of the real-world networks had negative  $\delta\text{-space}_{0.90}$  and  $\delta\text{-space}_{0.95}$  values. Nevertheless, there are six networks that continue to have a  $\delta\text{-space}_r$  value of 0.99 for all  $r$  values presented in Table 3. A closer look at these six networks reveals that the  $\lambda_{sp}$  (spectral radius ratio for node degree) values for all these six networks are less than 1.5. A further analysis of the  $\delta\text{-space}_{0.95}$  values of the real-world networks and the  $\lambda_{sp}$  values indicates that out of 25 networks that had a  $\lambda_{sp}$  value above 1.5, 22 networks (close to 90%) incurred negative values for  $\delta\text{-space}_{0.95}$ , whereas out of the remaining 23 networks (that had a  $\lambda_{sp}$  value below 1.5), only 7 networks (close to 30%) incurred negative values for  $\delta\text{-space}_{0.95}$ . This indicates a trend that networks with lower variation in node degree are more likely to have a positive  $\delta\text{-space}_r$  value (even for larger values for the threshold PCC  $r$ ), whereas networks with larger variation in node degree are more likely to incur negative values for  $\delta\text{-space}_r$ . In other words, for networks with larger variation in node degree, there is more likely not even a single  $\delta$  value for which we could expect a stronger correlation for DEC with both DEG and CLC simultaneously.

Table 3.  $\delta$ -space<sub>r</sub> and the Number of Decay Centrality Computations of Real-World Networks for Different Values of the Threshold PCC Values ( $r$ )

#	Net.	$\delta$ -space <sub>r</sub> / Number of Decay Centrality Computations									
		$r = 0.60$		$r = 0.70$		$r = 0.8$		$r = 0.90$		$r = 0.95$	
1	ADJ	0.99	1	0.99	1	0.64	1	0.25	3	0.04	6
2	AKN	0.99	1	0.95	1	0.59	1	0.26	3	0.05	6
3	JBN	0.99	1	0.99	1	0.75	1	0.36	3	0.09	5
4	CEN	0.99	1	0.57	1	0.28	3	0.05	4	-0.09	4
5	CLN	0.37	2	0.15	2	-0.11	2	-0.41	2	-0.59	2
6	CGD	0.65	2	0.36	2	0.04	5	-0.34	2	-0.58	2
7	CFN	0.99	1	0.99	1	0.89	1	0.55	1	0.22	3
8	DON	0.99	1	0.96	1	0.53	2	0.09	5	-0.19	3
9	DRN	0.99	1	0.51	2	0.01	7	-0.43	2	-0.64	2
10	DLN	0.99	1	0.99	1	0.99	1	0.58	1	0.24	3
11	ERD	0.23	4	0.00	4	-0.24	2	-0.53	2	-0.70	2
12	FMH	0.99	1	0.70	2	0.20	2	-0.29	2	-0.54	2
13	FHT	0.55	2	0.31	2	0.04	6	-0.28	2	-0.50	2
14	FTC	0.99	1	0.99	1	0.96	1	0.36	3	0.05	3
15	FON	0.17	4	0.08	4	0.00	9	-0.12	4	-0.23	3
16	CDF	0.99	1	0.99	1	0.99	1	0.99	1	0.99	1
17	GD96	0.49	2	0.29	3	0.09	3	-0.15	3	-0.33	3
18	MUN	0.03	2	0.06	4	-0.19	4	-0.47	2	-0.64	2
19	GLN	0.45	2	0.19	2	-0.09	5	-0.41	2	-0.60	2
20	HTN	0.99	1	0.99	1	0.99	1	0.99	1	0.99	1
21	HCN	0.30	2	0.09	5	-0.15	2	-0.43	2	-0.61	2
22	ISP	0.99	1	0.76	1	0.46	2	0.04	6	-0.24	3
22	KCN	0.99	1	0.99	1	0.63	2	0.19	3	-0.08	3
24	KFP	0.58	2	0.29	3	-0.01	9	-0.35	2	-0.56	2
25	LMN	0.99	1	0.99	1	0.62	1	0.21	3	-0.02	6
26	MDN	0.99	1	0.99	1	0.99	1	0.99	1	0.99	1
27	MTB	0.42	2	0.20	2	-0.04	6	-0.34	2	-0.54	2
28	MCE	0.99	1	0.99	1	0.99	1	0.99	1	0.99	1
29	MSJ	0.16	2	-0.03	5	-0.24	2	-0.50	2	-0.66	2
30	AFB	-0.03	5	-0.17	2	-0.32	2	-0.52	2	-0.67	2
31	MPN	0.99	1	0.99	1	0.99	1	0.59	2	0.18	3
32	MMN	0.99	1	0.99	1	0.66	2	0.11	2	-0.24	2
33	NSC	0.19	3	0.03	7	-0.16	4	-0.41	2	-0.60	2
34	PBN	0.88	2	0.54	2	0.23	2	-0.11	4	-0.35	2
35	PSN	0.99	1	0.99	1	0.99	1	0.99	1	0.60	1
36	PFN	0.99	1	0.99	1	0.99	1	0.49	2	0.13	3
37	SJN	0.99	1	0.68	2	0.34	3	0.02	3	-0.19	3
38	SDI	0.16	2	-0.14	2	-0.39	2	-0.61	2	-0.75	2
39	SPR	0.99	1	0.99	1	0.99	1	0.62	1	0.29	3
40	SWC	0.99	1	0.99	1	0.99	1	0.70	1	0.39	3
41	SSM	0.99	1	0.99	1	0.64	2	0.21	3	-0.05	3
42	TEN	0.93	1	0.59	2	0.29	3	-0.05	5	-0.28	3
43	TWF	0.36	2	0.14	3	-0.08	5	-0.38	2	-0.58	2
44	UKF	0.99	1	0.99	1	0.99	1	0.65	1	0.28	3
45	APN	0.99	1	0.79	1	0.5	1	0.18	3	-0.02	6
46	RHF	0.99	1	0.99	1	0.99	1	0.54	1	0.16	4
47	WSB	0.99	1	0.99	1	0.99	1	0.99	1	0.99	1
48	WTN	0.99	1	0.99	1	0.99	1	0.99	1	0.99	1

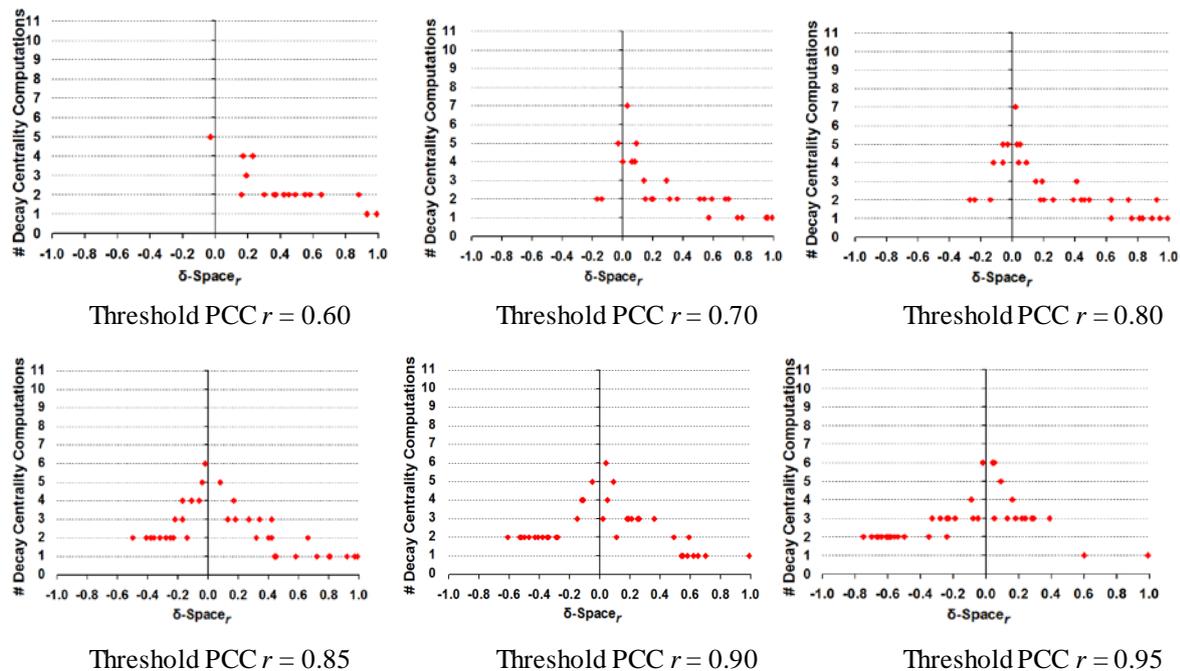
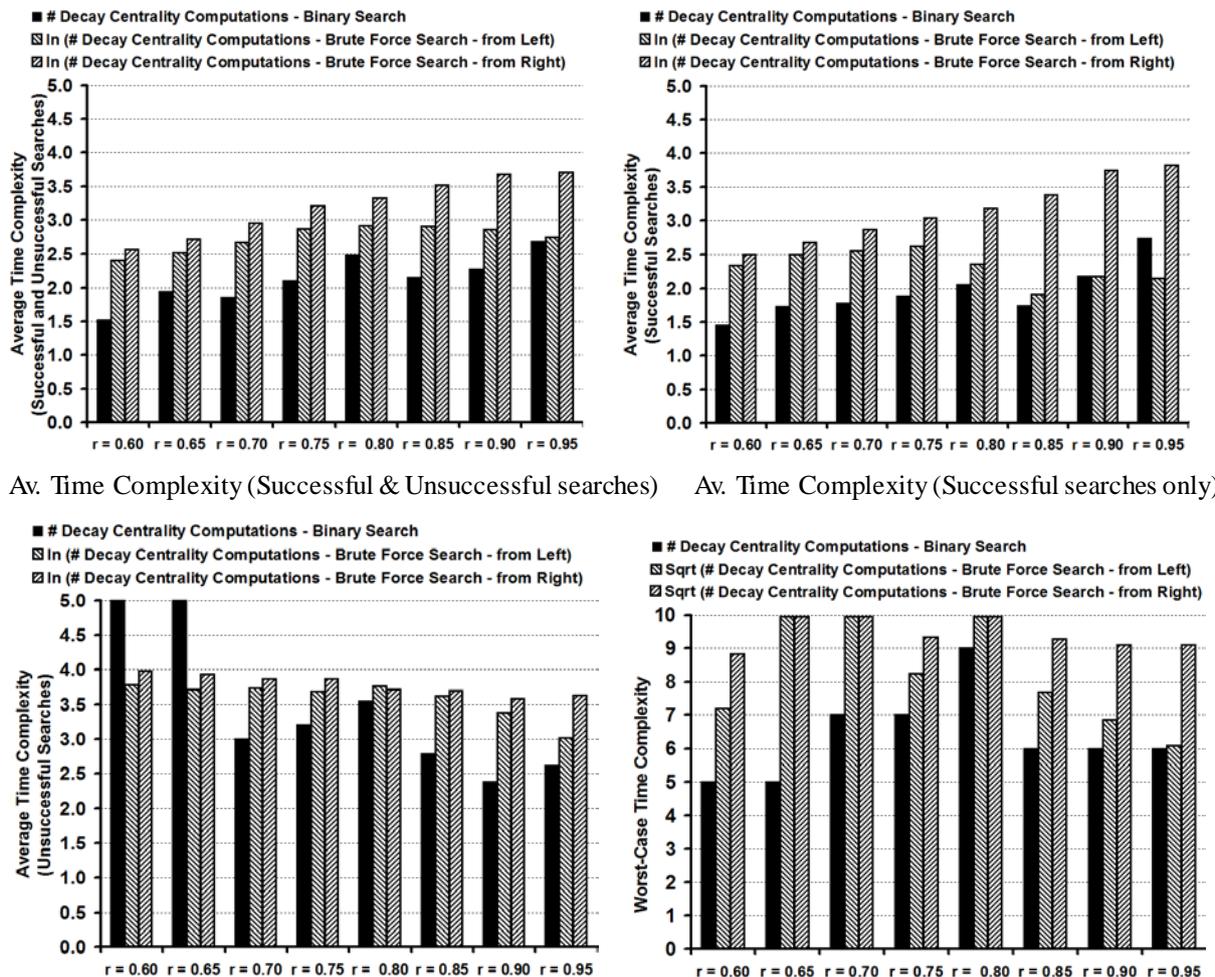


Figure 12. Binary Search Algorithm:  $\delta\text{-space}_r$  vs. # Decay Centrality Computations for Different Values of the Threshold Pearson's Correlation Coefficient ( $r$ )

Figure 12 presents the distribution of the number (#) of decay centrality computations vs. the  $\delta\text{-space}_r$  values of the real-world networks for different values of the threshold PCC ( $r$ ). We observe the # decay centrality computations to be lower for larger positive values of  $\delta\text{-space}_r$ , as well as for larger negative values of  $\delta\text{-space}_r$ . On the other hand, the # decay centrality computations is relatively larger for lower positive as well as lower negative values of  $\delta\text{-space}_r$ . This is because for larger values of  $\delta\text{-space}_r$ , it is more likely that the middle index of the binary search algorithm will soon correspond to a  $\delta$  value that falls within the range for  $\delta\text{-space}_r$ . If the  $\delta\text{-space}_r$  value is smaller, it takes relatively more iterations (and as a result a larger number of decay centrality computations) before the algorithm could identify a  $\delta$  value that falls within the range for  $\delta\text{-space}_r$ . Note that if  $\delta\text{-space}_r$  values are negative, it means there does not exist a  $\delta$ -value for which  $\text{PCC}(\text{DEC}_\delta, \text{DEG}) \geq r$  as well as  $\text{PCC}(\text{DEC}_\delta, \text{CLC}) \geq r$ ; this also means that there exists one or more  $\delta$  values for which  $\text{PCC}(\text{DEC}_\delta, \text{DEG}) < r$  as well as  $\text{PCC}(\text{DEC}_\delta, \text{DEG}) < r$ .

### 6.1 Binary Search vs. Brute Force Search

We now present a comparison of the time complexity (for both successful searches and unsuccessful searches considered together and separately) on the basis of the number of decay centrality computations incurred with the proposed binary search approach vs. brute force search approaches from left as well as from right of the  $\delta$ -search space. Under the brute force approach from left, for each real-world network and a given threshold PCC ( $r$ ), we iterate through the values of  $\delta$  from 0.01 to 0.99 (in this order), in increments of 0.01, and determine the smallest  $\delta$  value (if one exists) for which  $\text{PCC}(\text{DEC}_\delta, \text{DEG}) \geq r$  and  $\text{PCC}(\text{DEC}_\delta, \text{CLC}) \geq r$ . Under the brute force approach from right, we iterate through the values of  $\delta$  from 0.99 to 0.01 (in this order), in decrements of 0.01, and determine the largest  $\delta$  value (if one exists) for which  $\text{PCC}(\text{DEC}_\delta, \text{DEG}) \geq r$  and  $\text{PCC}(\text{DEC}_\delta, \text{CLC}) \geq r$ .



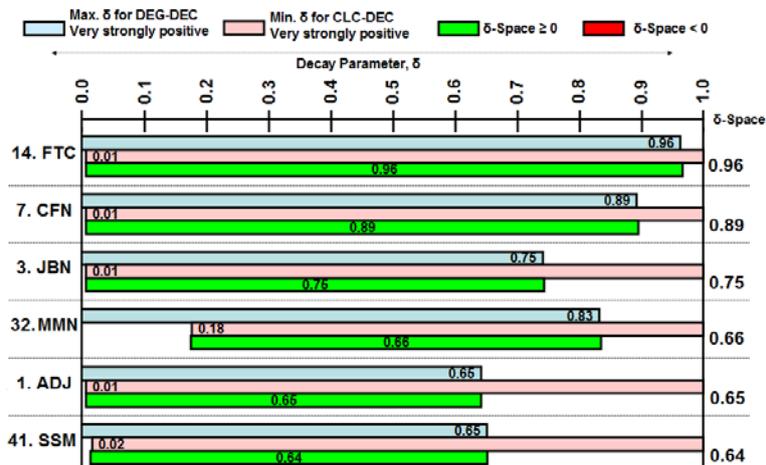
Av. Time Complexity (Successful & Unsuccessful searches)

Av. Time Complexity (Successful searches only)

Av. Time Complexity (Unsuccessful searches only)

Worst-Case Time Complexity

Figure 13. Comparison of the Time Complexity (# Decay Centrality Computations) of the Binary Search Algorithm vs. Brute Force Search for Different Values of Threshold Pearson's Correlation Coefficient ( $r$ )



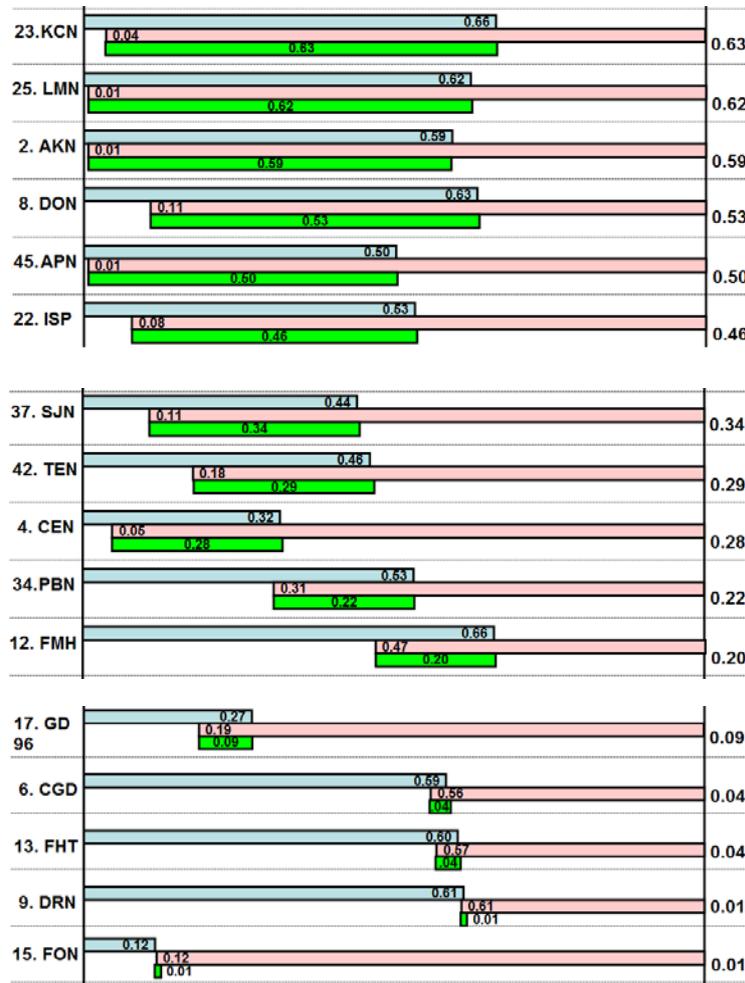


Figure 14. Real-World Networks in the Decreasing Order of  $\delta\text{-space}_{r=vs+}$  Values ( $0 < \delta\text{-space}_{r=vs+} < 0.99$ )

To be fair to the brute force approaches, we adopt the same condition (used for the binary search approach) to terminate the search a priori (instead of searching through the entire  $\delta$ -search space) if we encounter a  $\delta$  value for which both  $PCC(DEC_\delta, DEG) < r$  and  $PCC(DEC_\delta, CLC) < r$ . When such a  $\delta$  value is encountered for a given threshold  $PCC(r)$  for a real-world network, it implies the  $\delta\text{-space}_r$  for the real-world network is negative and the search for a  $\delta$  to satisfy the threshold  $PCC$  value ( $r$ ) would be unsuccessful. The number of decay centrality iterations incurred until encountering the smallest  $\delta$  value (if proceeded from the left) or the largest  $\delta$  value (if proceeded from the right) that satisfies the condition for a successful search or the  $\delta$  value that satisfies the condition for an unsuccessful search is recorded. For the binary search and the two variants of brute force search algorithms and for each value of the threshold  $PCC(r)$ : from 0.60 to 0.95, in increments of 0.05, we determine the average number of decay centrality iterations (averaged over all the 48 real-world networks) for successful search and unsuccessful search considered together and considered separately as well as the worst-case number of decay centrality iterations (the maximum of the number of decay centrality iterations incurred among the 48 real-world networks).

We observe the binary search algorithm to comprehensively outperform the brute force search methods (from left and from right). In Figure 13, the average-case number of decay centrality computations (when the successful and unsuccessful searches are considered together and when considered alone) for the binary search algorithm are very much comparable (or even lower in most of the cases) to the logarithm of the number of decay centrality computations incurred with the brute force search methods. The worst-case number of decay centrality computations incurred with the binary search method is appreciably lower than the square root of the worst-case number of decay centrality computations incurred with the brute force search methods.

It is to be noted that the # decay centrality computations incurred with the binary search method (for both

average and worst-case) are not plotted in logarithm scale. Hence, even if the bars for the binary search method appear to be taller than the bars for the brute force search methods (this happens only in a few cases:  $r = 0.60$  and  $0.65$  for average time complexity - unsuccessful searches only, and  $r = 0.95$  for average time complexity - successful searches only), the raw values for the average # decay centrality computations incurred with the binary search method for all the three scenarios (successful searches only, unsuccessful searches only, and successful and unsuccessful searches together) are appreciably smaller than the raw values for the average # decay centrality computations incurred with the brute force search methods.

We observe the brute force approach of proceeding from the left (i.e., starting from  $\delta = 0.01$ , for which the PCC of DEC and DEG is observed to be the largest and is 0.95 or above for all the real-world networks) to be relatively better than the brute force approach of proceeding from the right (i.e., starting from  $\delta = 0.99$ , for which the PCC of DEC and CLC is observed to be the largest and is 0.95 or above for all the real-world networks). This could be attributed to the degree centrality being a bottleneck metric vis-a-vis the closeness centrality metric when evaluated on the basis of the range of  $\delta$  values for which the PCC of decay centrality with the two metrics is above a certain threshold ( $r$ ).

We justify this attribution using an example scenario (threshold PCC  $r$  value of 0.80). Figure 14 presents the real-world networks in the decreasing order of the  $\delta\text{-space}_{r=0.80}$  values, wherein  $0 < \delta\text{-space}_{r=0.80} < 0.99$  (there are 22 such real-world networks out of the total of 48). We observe  $\delta_{PCC(DEC,CLC) \geq 0.80}^{\min}$  to be less than 0.20 for 17 of the 22 real-world networks (implying the closeness centrality metric provided the opportunity for  $\delta\text{-space}_{r=0.80}$  to be 0.80 or above for 17 of the 22 real-world networks for which  $\delta\text{-space}_{r=0.80} < 0.99$ ) whereas  $\delta_{PCC(DEC,DEG) \geq 0.80}^{\max}$  is greater than 0.80 for only 3 of the 22 real-world networks that are of interest in this case (implying the degree centrality metric provided the opportunity for  $\delta\text{-space}_{r=0.80}$  to be 0.80 or above for only 3 of the 22 real-world networks for which  $\delta\text{-space}_{r=0.80} < 0.99$ ). Hence, it is obvious that the lower  $\delta\text{-space}_{r=0.80}$  values for the 22 real-world networks is due to the reduced range of  $\delta$  values for which  $PCC(DEC_{\delta}, DEG) \geq 0.80$  vis-a-vis  $PCC(DEC_{\delta}, CLC) \geq 0.80$ .

6.2 Determination of the Maximum Threshold PCC Value ( $r_{max}$ ) for a Positive  $\delta\text{-Space}_{max}$

From Table 3, it is obvious that the  $\delta\text{-space}_r$  values turn negative for one or more real-world networks as the value for the threshold PCC ( $r$ ) increases. However, Table 3 does not show the exact threshold PCC value above which  $\delta\text{-space}$  is no longer positive. We now illustrate the use of the binary search algorithm to determine the maximum Threshold PCC value ( $r_{max}$ ) such that  $\delta\text{-space}_{r_{max}}$  is positive. The procedure to determine the  $r_{max}$  value is described in the last paragraph of Section 4. Since the  $\delta\text{-space}_r$  values turn negative as the threshold PCC approaches 1 (rather than 0), we start with a tentative threshold PCC ( $r_{max}$ ) value of 1.0 and decrement the value by 0.01 until we come across a  $r$  value for which the binary search algorithm finds the existence of a positive  $\delta\text{-space}_r$ . The corresponding  $\delta$  value (that belongs to the positive  $\delta\text{-space}_r$  and is referred to as  $\delta_{r_{max}}$ ) determined by the binary search algorithm is the  $\delta$  value at which the distributions for the  $PCC(DEC_{\delta}, DEG)$  and  $PCC(DEC_{\delta}, CLC)$  intersect.

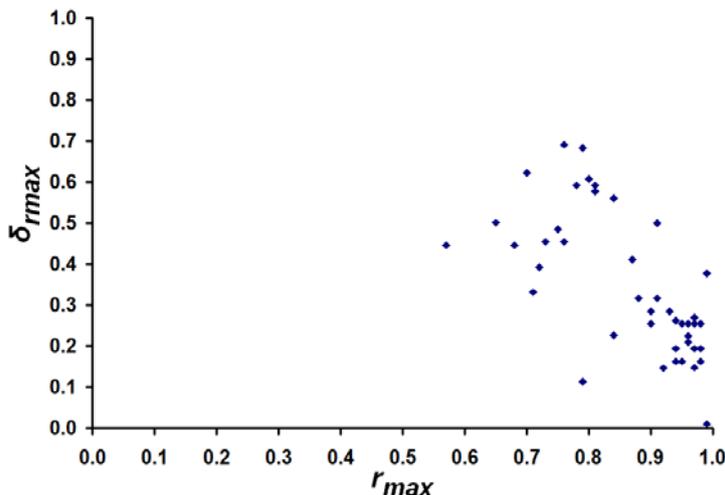


Figure 15. Distribution of the  $r_{max}$  and  $\delta_{r_{max}}$  values for Real-World Networks

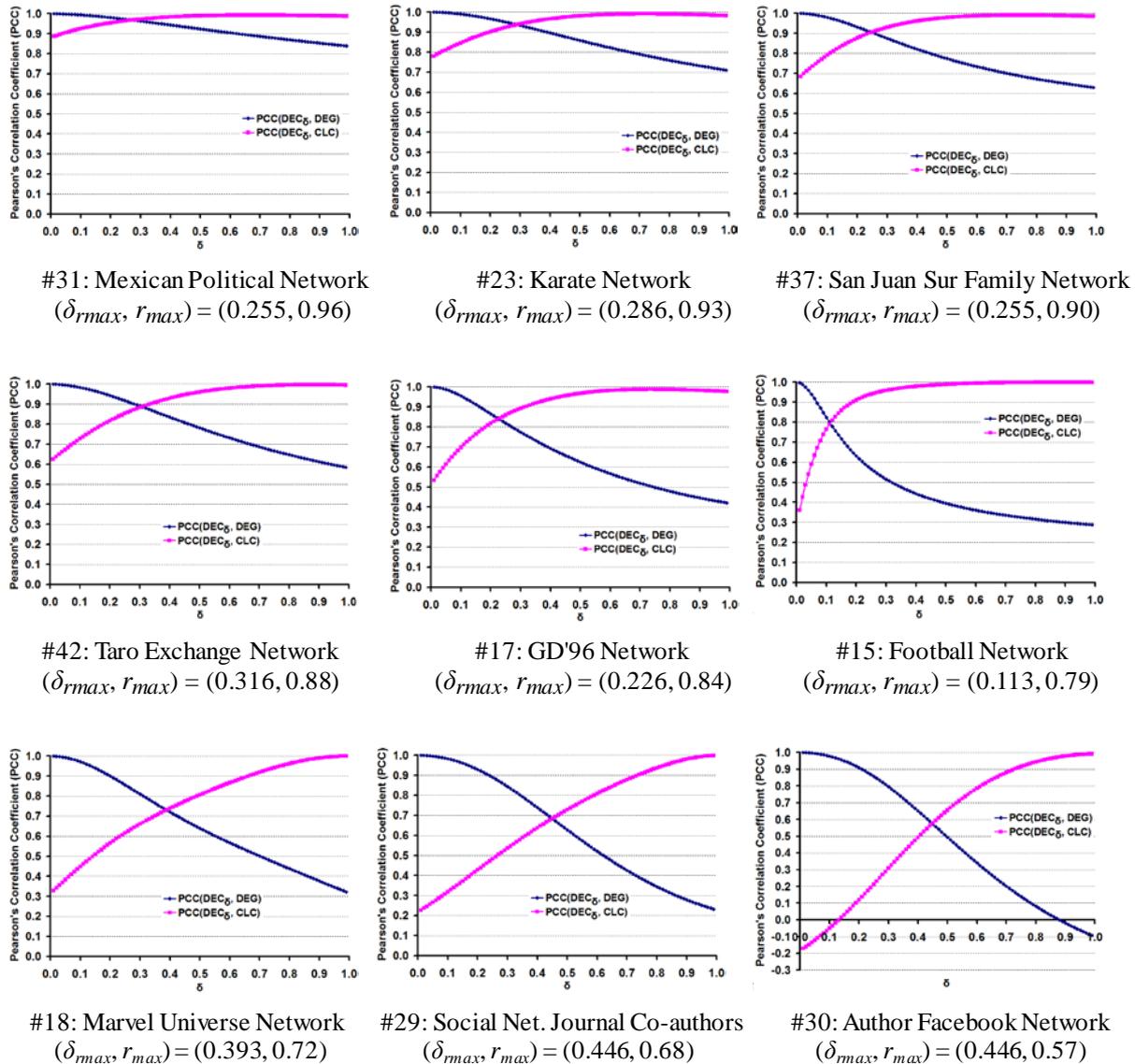


Figure 16. Samples of the PCC Distributions for DEC vs. DEG and DEC vs. CLC and their Intersection for Real-World Networks in the Decreasing Order of the  $r_{max}$  Values

Table 4 presents the maximum threshold PCC  $r_{max}$  values and the corresponding  $\delta_{r_{max}}$  values for the 48 real-world networks. Figure 15 presents a distribution of the  $r_{max}$  and the corresponding the  $\delta_{r_{max}}$  values; the trend one could infer is that real-world networks with a very large  $r_{max}$  value are more likely to have a low  $\delta_{r_{max}}$  value and real-world networks with a moderate  $r_{max}$  value are more likely to have a moderately large  $\delta_{r_{max}}$  value. The median of the  $r_{max}$  and  $\delta_{r_{max}}$  values are 0.915 and 0.266 respectively. Figure 16 presents the  $PCC(DEC_{\delta}, DEG)$  vs.  $PCC(DEC_{\delta}, CLC)$  distributions for some of the real-world networks (in the decreasing order of the  $r_{max}$  values) to illustrate the intersection of the two distributions at the  $\delta_{r_{max}}$  value. Listed below each of the sub figures of Figure 16 are the  $(\delta_{r_{max}}, r_{max})$  values observed for the real-world network in consideration.

Table 4. Maximum Threshold PCC ( $r_{max}$ ) Values and Corresponding  $\delta_{r_{max}}$  Values for the Real-World Networks

#	Net.	$r_{max}$	$\delta_{r_{max}}$	#	Net.	$r_{max}$	$\delta_{r_{max}}$	#	Net.	$r_{max}$	$\delta_{r_{max}}$
1	ADJ	0.95	0.163	17	GD96	0.72	0.393	33	NSC	0.71	0.332
2	AKN	0.95	0.163	18	MUN	0.76	0.454	34	PBN	0.87	0.412
3	JBN	0.96	0.209	19	GLN	0.99	0.010	35	PSN	0.98	0.255
4	CEN	0.92	0.147	20	HTN	0.73	0.454	36	PFN	0.96	0.224

5	CLN	0.75	0.485	21	HCN	0.90	0.286	37	SJN	0.90	0.255
6	CGD	0.81	0.577	22	ISP	0.93	0.286	38	SDI	0.65	0.501
7	CFN	0.97	0.270	22	KCN	0.79	0.684	39	SPR	0.98	0.163
8	DON	0.91	0.316	24	KFP	0.94	0.194	40	SWC	0.98	0.194
9	DRN	0.80	0.607	25	LMN	0.99	0.010	41	SSM	0.94	0.263
10	DLN	0.97	0.255	26	MDN	0.78	0.592	42	TEN	0.88	0.316
11	ERD	0.70	0.623	27	MTB	0.99	0.010	43	TWF	0.76	0.691
12	FMH	0.84	0.561	28	MCE	0.68	0.446	44	UKF	0.97	0.194
13	FHT	0.81	0.592	29	MSJ	0.57	0.446	45	APN	0.94	0.163
14	FTC	0.95	0.255	30	AFB	0.96	0.255	46	RHF	0.97	0.148
15	FON	0.79	0.113	31	MPN	0.91	0.500	47	WSB	0.99	0.378
16	CDF	0.99	0.010	32	MMN	0.72	0.393	48	WTN	0.99	0.010

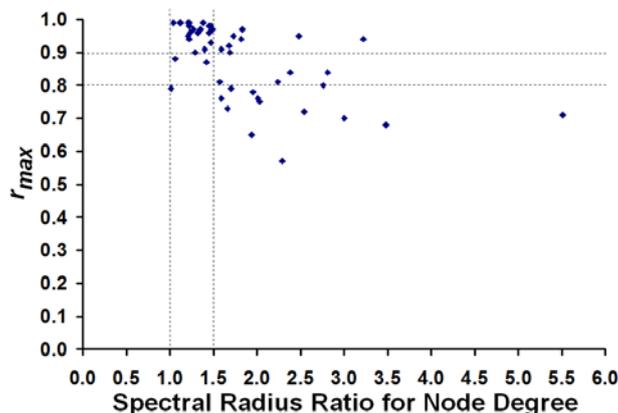


Figure 17. Spectral Radius Ratio for Node Degree vs. Maximum Threshold PCC ( $r_{max}$ ) for Real-World Networks

The sub figures of Figure 16 also stand as testimony to our earlier statement about the results of our initial correlation study that the  $PCC(DEC_{\delta}, DEG)$  monotonically decreases with increase in  $\delta$  and  $PCC(DEC_{\delta}, CLC)$  monotonically increases with increase in  $\delta$ . It is this phenomenon that forms the backbone of our binary search algorithm and enabled us to initially setup the left and right indexes to correspond to the  $\delta$  values at which we observe the max PCC value with DEG and CLC respectively, and later move these indexes towards each other by maintaining the invariant that the left index always corresponds to a  $\delta$  value at which the  $PCC(DEC_{\delta}, DEG)$  is greater than or equal to the threshold PCC ( $r$ ) and likewise the right index always corresponds to a  $\delta$  value at which the  $PCC(DEC_{\delta}, CLC)$  is greater than or equal to the threshold PCC ( $r$ ). The algorithm seeks to find a middle index (if  $\delta\text{-space}_r$  is positive), which corresponds to the average of the  $\delta$  values represented by the left index and right index, such that the  $PCC(DEC_{\text{middle index}}, DEG)$  and  $PCC(DEC_{\text{middle index}}, CLC)$  are both greater than or equal to the threshold PCC ( $r$ ). If the algorithm comes across a middle index for which both  $PCC(DEC_{\text{middle index}}, DEG)$  and  $PCC(DEC_{\text{middle index}}, CLC)$  are less than the threshold PCC ( $r$ ), it is guaranteed (due to the monotonically non-increasing and non-decreasing trend of the PCC values) that the  $\delta\text{-space}_r$  for the real-world network is negative.

Like in the case of  $\delta\text{-space}_r$ , we also observe networks with larger variation in node degree to incur lower values for  $r_{max}$  and networks with low variation in node degree to incur larger  $r_{max}$  values. Of the 27 real-world networks that had  $r_{max}$  values of 0.9 or above, 19 of these networks had  $\lambda_{sp}$  (spectral radius ratio for node degree) values less than 1.5. On the other hand, 12 of the 13 real-world networks with  $r_{max}$  values less than 0.8 had  $\lambda_{sp}$  values above 1.5. Figure 17 plots the distribution of the  $r_{max}$  vs.  $\lambda_{sp}$  values for real-world networks.

### 6.3 Correlation between the Maximum Threshold PCC Value ( $r_{max}$ ) and the Pearson's Correlation between Degree and Closeness Centrality Metrics

As part of further analysis, we analyzed the correlation between the  $r_{max}$  values observed for the real-world networks and the Pearson's correlation coefficient between DEG and CLC. We observe a very strong positive correlation between the  $PCC(DEG, CLC)$  and the  $r_{max}$  values. The regression equation is shown below (equation 4); the  $R^2$  for this straight line fit (shown in Figure 18) is 0.9485 and the Standard Error of the Residuals (SER) is

0.025, a significantly smaller value given that the range for  $r_{max}$  is from -1 to 1. Figure 19 presents the distribution of the actual  $r_{max}$  values vs. the  $r_{max}$  values predicted using the actual values of the PCC(DEG, CLC) and the regression equation (4); we observe the data points to lie close to the diagonal line, justifying the smaller SER value for the prediction.

$$\text{Predicted } r_{max} = 0.3792 * \text{PCC}(\text{DEG}, \text{CLC}) + 0.626 \tag{4}$$

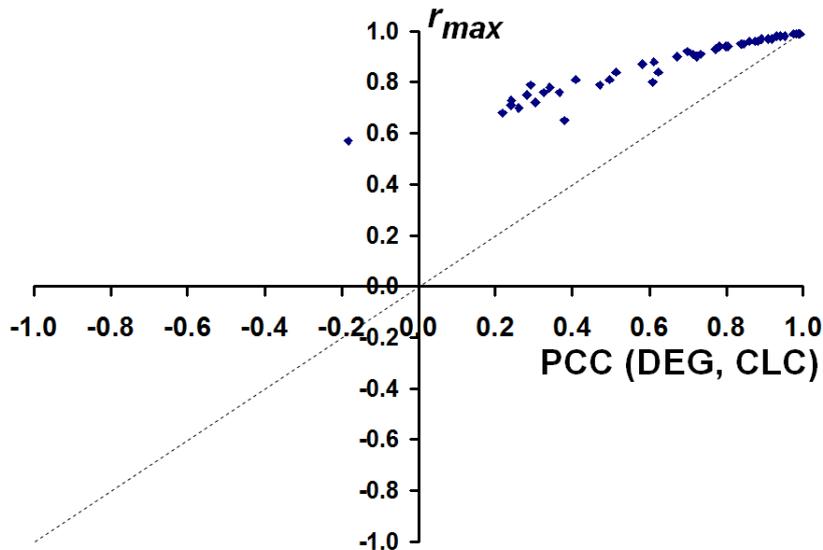


Figure 18. PCC (DEG, CLC) vs. the Maximum Threshold PCC ( $r_{max}$ ) Value for Real-World Networks

Though there is a very strong linear correlation, we observe (from Figure 17) the  $r_{max}$  value for a real-world network to be typically much larger than the PCC(DEG, CLC) value for the network. The median of the PCC (DEG, CLC) values is 0.728, while the median of the  $r_{max}$  values is 0.915. Note that, as per the ordinal scale proposed by Evans (1995), 0.8 is typically the minimum correlation coefficient value expected for two metrics to be considered to exhibit a very strongly positive correlation. We observe 35 of the 48 real-world networks (i.e., more than 2/3rds of the networks) to have a  $r_{max}$  value of 0.8 or above. Thus, though the PCC(DEG, CLC) for a real-world network might be low, we observe that there exist at least one value of  $\delta$  ( $\delta_{max}$  that could be efficiently found by our binary search algorithm) for which we could simultaneously find a relatively stronger correlation between DEG and DEC as well as between CLC and DEC.

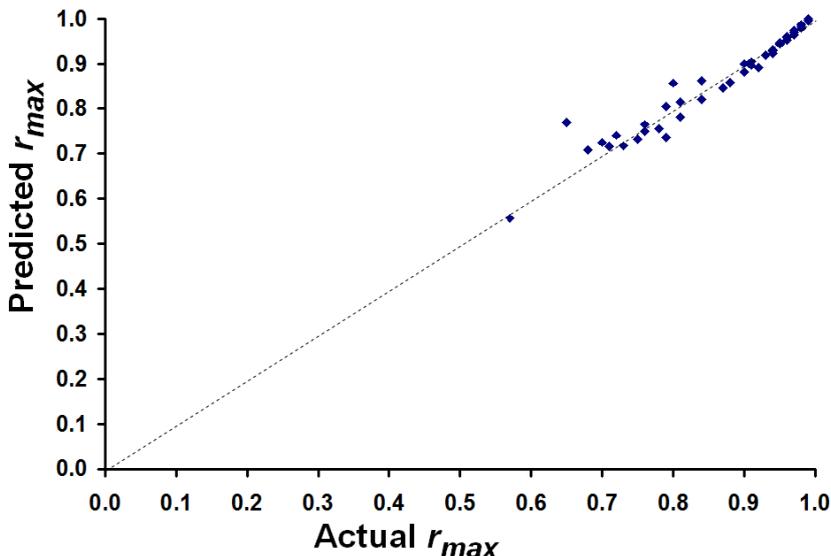


Figure 19. Distribution of Actual vs. Predicted  $r_{max}$  Values for the Real-World Networks

## 7. Related Work and Our Contributions

Decay centrality has not been explored much in the literature for complex network analysis. To the best of our knowledge, ours is the first work to conduct a correlation study focusing on decay centrality. Most of the work (e.g., Li et. al., 2015, Meghanathan, 2015) on correlation studies (involving centrality metrics) were focused on the commonly studied centrality metrics such as the neighborhood-based degree centrality and eigenvector centrality (Bonacich, 1987) and shortest path-based betweenness centrality (Freeman, 1977) and closeness centrality. The objective of such correlation studies has been typically to identify computationally-light alternatives (like DEG and its derivatives; Meghanathan, 2017) for computationally-heavy metrics (such as EVC and BWC) for both real-world networks and simulated networks of theoretical models (Renyi, 1959; Barabasi & Albert, 1999). The focus of our paper is different from such typical correlation studies in the literature. We seek to explore the trend of change in the correlation coefficients between a parameter-driven centrality metric (whose values for a node change for different values of the decay parameter) and the degree and closeness centrality metrics whose values are not parameter-driven and remain the same for a particular network.

The most related work to our work is a recent study (Tsakas, 2016) on random networks (Renyi, 1959) for which a single threshold value of the decay parameter (referred here as  $\delta_{thresh}$ ) was observed to exist (for a particular operating condition) such that nodes with high degree centrality also had a high decay centrality computed for  $\delta$  values less than  $\delta_{thresh}$  and nodes with high closeness centrality also had a high decay centrality computed for  $\delta$  values above  $\delta_{thresh}$ . It was observed by Tsakas (2016) that for random networks: nodes with the largest values for degree centrality and closeness centrality are more likely to be nodes that also incur the largest values for decay centrality for almost all values of  $\delta$ . In addition, nodes that had the largest decay centrality for a certain value of  $\delta$  are more likely to be part of the set of nodes that had the largest degree centrality or the largest closeness centrality. The likelihood of all of the above was studied using multinomial logistic regression (Greene, 2011). Most of the other works (e.g., Chatterjee & Dutta, 2015; Kang et. al., 2012) on decay centrality metric have focused on exploring its suitability for diffusion in socio-economic networks with regards to selecting the seed nodes that could effectively propagate information about a product to putative customers. Nodes that are themselves central and connected to other central nodes (via direct links or shorter paths) in the network are typically preferred for such "agent" roles (Tsakas, 2016; Chatterjee & Dutta, 2015). The use of decay centrality vis-a-vis diffusion centrality (Kang et. al., 2012) and eigenvector centrality (Ide et. al., 2014; Banerjee et. al., 2013) to identify such "agent" nodes for diffusion has been explored in the literature.

Our paper differs from all of the above work and is innovative on the following lines: We analyze real-world networks rather than the simulated random networks. We use the Pearson's correlation measure to study the correlation between the actual centrality values rather than multinomial logistic regression (Greene, 2011) to study the sets of vertices that had the largest values of centrality. We have unearthed the trend (not known until now) that the Pearson's correlation coefficient between decay centrality and degree centrality decreases with increase in the value of the decay parameter  $\delta$  and that the Pearson's correlation coefficient between decay centrality and closeness centrality decreases with decrease in  $\delta$ . We have developed an efficient binary search algorithm that makes use of the above phenomenon to determine the existence (or the lack of it) of one or more  $\delta$  values (collectively referred to as  $\delta$ -space<sub>r</sub>) for which  $PCC(DEC_{\delta}, DEG) \geq r$  and  $PCC(DEC_{\delta}, CLC) \geq r$  for a threshold PCC ( $r$ ). We also demonstrate the use of the binary search algorithm to determine the maximum threshold PCC ( $r_{max}$ ) value that could be observed between DEC-DEG as well as between DEC-CLC for a real-world network. We observe this  $r_{max}$  value to be appreciably larger than  $PCC(DEC, CLC)$  for most of the real-world networks, and also show that it could be accurately predicted using the latter. One could thus run the binary search algorithm in the vicinity of the predicted  $r_{max}$  value for a real-world networks and determine a  $\delta$  value for which we observe the maximum threshold PCC between DEC and DEG as well as between DEC and CLC.

## 8. Conclusions

Our contributions in this paper are as follows: For each of the 48 real-world networks (of diverse degree distributions) analyzed in this paper, we observe the Pearson's Correlation Coefficient (PCC) between degree centrality (DEG) and decay centrality (DEC) to monotonically decrease with increase in the decay parameter ( $\delta$ ), and the PCC between closeness centrality (CLC) and decay centrality to monotonically increase with increase in  $\delta$ . We have explored this phenomenon and proposed a binary search algorithm that could be used (for a given threshold PCC  $r$ ) to determine the existence of a positive  $\delta$ -space<sub>r</sub> (or the absence of the same) comprising of one or more  $\delta$  values for which  $PCC(DEC_{\delta}, DEG) \geq r$  and  $PCC(DEC_{\delta}, CLC) \geq r$ . In addition, we show the use of the binary search algorithm to determine the maximum threshold PCC ( $r_{max}$ ) value for a real-world network and the prediction of the same using  $PCC(DEC, CLC)$ . The  $r_{max}$  value for a real-world network would be a measure of

the extent to which the degree centrality or closeness centrality metrics could serve as alternatives to the decay centrality metric and vice-versa. If the predicted  $r_{max}$  value for a real-world network is high, then one could run the binary search algorithm in the vicinity of the  $r_{max}$  value to determine a value of  $\delta$  ( $\delta_{max}$ ) for which  $PCC(DEC_{\delta_{max}}, DEG) = r_{max}$  and  $PCC(DEC_{\delta_{max}}, CLC) = r_{max}$ . As vertices with large decay centrality are preferable for diffusion, our approach of determining the  $r_{max}$  value and the corresponding  $\delta_{max}$  value for a real-world network using the proposed binary search algorithm could bring significant savings in the process of exploring a suitable  $\delta$  value for which DEC exhibits the largest correlation coefficient value with both DEG and CLC.

### Acknowledgments

The research is financed by the NASA EPSCoR sub award (#: NNX14AN38A) from University of Mississippi.

### References

- Banerjee, A., Chandrasekhar, G., Duflo, E., & Jackson, M. O. (2013). The Diffusion of Microfinance. *Science*, 341(6144), 1-49. <https://doi.org/10.1126/science.1236498>
- Barabasi, A. L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286(5439), 509-512. <https://doi.org/10.1126/science.286.5439.509>
- Batagelj, V., & Mrvar, A. (2006). *Pajek Datasets*. Retrieved from <http://vlado.fmf.uni-lj.si/pub/networks/data/>
- Bernard, H. R., Killworth, P. D., & Sailer, L. (1980). Informant Accuracy in Social Network Data IV: A Comparison of Clique-level Structure in Behavioral and Cognitive Network Data. *Social Networks*, 2(3), 191-218. [https://doi.org/10.1016/0378-8733\(79\)90014-5](https://doi.org/10.1016/0378-8733(79)90014-5)
- Biedl, T., & Franz, B. J. (2001). *Graph-Drawing Contest Report*. Paper presented at the 9th International Symposium on Graph Drawing, Vienna, Austria. [https://doi.org/10.1007/3-540-45848-4\\_60](https://doi.org/10.1007/3-540-45848-4_60).
- Bonacich, P. (1987). Power and Centrality: A Family of Measures. *American Journal of Sociology*, 92(5), 1170-1182. <https://doi.org/10.1086/228631>
- Chatterjee, K., & Dutta, B. (2015). Credibility and Strategic Learning in Networks. *International Economic Review*, 57(3), 759-786. <https://doi.org/10.1111/iere.12175>
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms*. (3<sup>rd</sup> ed.) MIT Press.
- Cross, R. L., Parker, A., & Cross, R. (2004). *The Hidden Power of Social Networks: Understanding How Work Really Gets Done in Organizations*. (1<sup>st</sup> ed.) Harvard Business Review Press.
- Evans, J. D. (1995). *Straightforward Statistics for the Behavioral Sciences*. (1<sup>st</sup> ed.) Brooks Cole Publishing Company.
- Freeman, L. (1977). A Set of Measures of Centrality based on Betweenness. *Sociometry*, 40(1), 35-41. <https://doi.org/10.2307/3033543>
- Freeman, L. (1979). Centrality in Social Networks: Conceptual Clarification. *Social Networks*, 1(3), 215-239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
- Freeman, L., Freeman, S. C., & Michaelson, A. G. (1989). How Humans See Social Groups: A Test of the Sailer-Gaulin Models. *Journal of Quantitative Anthropology*, 1, 229-238
- Freeman, L., Webster, C. M., & Kirke, D. M. (1998). Exploring Social Structure using Dynamic Three-Dimensional Color Images. *Social Networks*, 20(2), 109-118. [https://doi.org/10.1016/S0378-8733\(97\)00016-6](https://doi.org/10.1016/S0378-8733(97)00016-6)
- Geiser, P., & Danon, L. (2003). Community Structure in Jazz. *Advances in Complex Systems*, 6(4), 563-573. <https://doi.org/10.1142/S0219525903001067>
- Gemmetto, V., Barrat, A., & Cattuto, C. (2014). Mitigation of Infectious Disease at School: Targeted Class Closure vs. School Closure. *BMC Infectious Diseases*, 14(695), 1-10. <https://doi.org/10.1186/s12879-014-0695-9>
- Gil-Mendieta, J., & Schmidt, S. (1996). The Political Network in Mexico. *Social Networks*, 18(4), 355-381. [https://doi.org/10.1016/0378-8733\(95\)00281-2](https://doi.org/10.1016/0378-8733(95)00281-2)
- Girvan, M., & Newman, M. (2002). Community Structure in Social and Biological Networks. *Proceedings of the National Academy of Sciences USA*, 99(12), 7821-7826. <https://doi.org/10.1073/pnas.122653799>

- Gleiser, P. M. (2007). How to become a Superhero. *Journal of Statistical Mechanics: Theory and Experiments*, P09020. <https://doi.org/10.1088/1742-5468/2007/09/P09020>
- Greene, W. H. (2011). *Econometric Analysis*. (7<sup>th</sup> ed.) Pearson.
- Grimmer, J. (2010). A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases. *Political Analysis*, 18(1), 1-35. <https://doi.org/10.1093/pan/mpp034>
- Hayes, B. (2006). Connecting the Dots. *American Scientist*, 94(5), 400-404. <https://doi.org/10.1511/2006.61.3495>
- Hummon, N. P., Doreian, P., & Freeman, L. C. (1990). Analyzing the Structure of the Centrality-Productivity Literature created between 1948 and 1979. *Science Communication*, 11(4), 459-480. <https://doi.org/10.1177/107554709001100405>
- Ide, K., Namatame, A., Ponnambalam, L., Xiuju, F., & Goh, R. S. M. (2014). A New Centrality Measure for Probabilistic Diffusion in Network. *Advances in Computer Science: An International Journal*, 3(5), 115-121, <https://doi.org/10.1007/s40595-016-0073-1>
- Isella, L., Stehle, J., Barrat, A., Cattuto, C., Pinton, J. F., & Van den Broeck, W. (2011). What's in a Crowd? Analysis of Face-to-Face Behavioral Networks. *Journal of Theoretical Biology*, 271(1), 161-180. <https://doi.org/10.1016/j.jtbi.2010.11.033>
- Jackson, M. O. (2010). *Social and Economic Networks*. (1<sup>st</sup> ed.) Princeton University Press.
- Kang, C., Molinaro, C., Kraus, S., Shavitt, Y., & Subrahmanian, V. S. (2012). *Diffusion Centrality in Social Networks*. Paper presented at the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Istanbul, Turkey. <https://doi.org/10.1109/ASONAM.2012.95>
- Knuth, D. E. (1993). *The Stanford GraphBase: A Platform for Combinatorial Computing*. (1<sup>st</sup> ed.) Addison-Wesley.
- Krackhardt, D. (1999). The ties that torture: Simmelian tie analysis in organizations. *Research in the Sociology of Organizations*, 16(1), 183-210.
- Krebs, V. (2003). Proxy Networks: Analyzing One Network to Reveal Another. *Bulletin de Méthodologie Sociologique*, 79, 61-70.
- Lay, D. C., Lay, S. R., and McDonald, J. J. (2015). *Linear Algebra and its Applications*, (5<sup>th</sup> ed.) Pearson.
- Lee, J. S. (2004). Generating Networks of Illegal Drug Users using Large Samples of Partial Ego-Network Data. *Intelligence and Security Informatics, Lecture Notes in Computer Science*, 3073, 390-402. [https://doi.org/10.1007/978-3-540-25952-7\\_29](https://doi.org/10.1007/978-3-540-25952-7_29)
- Li, C., Li, Q., Van Mieghem, P., Stanley, H. E., & Wang, H. (2015). Correlation between Centrality Metrics and their Application to the Opinion Model. *The European Physical Journal B*, 88(35), 1-13.
- Loomis, C. P., Morales, J. O., Clifford, R. A., & Leonard, O. E. (1953). *Turrialba Social Systems and the Introduction of Change*. (1<sup>st</sup> ed.) The Free Press.
- Lusseau, D., Schneider, K., Boisseau, O. J., Hasse, P., Slooten, E., & Dawson, S. M. (2003). The Bottleneck Dolphin Community of Doubtful Sound Features a Large Proportion of Long-lasting Associations. *Behavioral Ecology and Sociobiology*, 54(3), 396-405. <https://doi.org/10.1007/s00265-003-0651-y>
- MacRae, D. (1960). Direct Factor Analysis of Sociometric Data. *Sociometry*, 23(4), 360-371.
- McCarty, C., & Freeman, L. (2008). *Network Datasets*. Retrieved from <http://moreno.ss.uci.edu/data.html>.
- Meghanathan, N. (2014). *Spectral Radius as a Measure of Variation in Node Degree for Complex Network Graphs*. Paper presented at the 3rd International Conference on Digital Contents and Applications, Hainan, China. <https://doi.org/10.1109/UNESST.2014.8>
- Meghanathan, N. (2015). Correlation Coefficient Analysis of Centrality Metrics for Complex Network Graphs. Paper presented at the 4th Computer Science Online Conference. [https://doi.org/10.1007/978-3-319-18503-3\\_2](https://doi.org/10.1007/978-3-319-18503-3_2)
- Meghanathan, N. (2017). A Computationally-Lightweight and Localized Centrality Metric in lieu of Betweenness Centrality for Complex Network Analysis. *Springer Vietnam Journal of Computer Science*, 4(1), 23-38, <https://doi.org/10.1007/s40595-016-0073-1>
- Michael, J. H. (1997). Labor Dispute Reconciliation in a Forest Products Manufacturing Facility. *Forest*

*Products Journal*, 47(11-12), 41-45.

Moreno, J. L. (1960). *The Sociometry Reader*. (1<sup>st</sup> ed.) The Free Press.

Nepusz, T., Petroczi, A., Negyessy, L., & Bazso, F. (2008). Fuzzy Communities and the Concept of Bridgeness in Complex Networks. *Physical Review E*, 77(1), 016107.

Newman, M. (2006). Finding Community Structure in Networks using the Eigenvectors of Matrices. *Physical Review E*, 74(3), 036104. <https://doi.org/10.1103/PhysRevE.74.036104>

Nooy, W. (1999). A Literary Playground: Literary Criticism and Balance Theory. *Poetics*, 26(5-6), 385-404. [https://doi.org/10.1016/S0304-422X\(99\)00009-1](https://doi.org/10.1016/S0304-422X(99)00009-1)

Pearson, M., & Michell, L. (2000). Smoke Rings: Social Network Analysis of Friendship Groups, Smoking and Drug-taking. *Drugs: Education, Prevention and Policy*, 7(1), 21-37. <https://doi.org/10.1080/dep.7.1.21.37>

Renyi, E. (1959). On Random Graphs I. *Publicationes Mathematicae*, 6, 290-297.

Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J., Tabor, J., Beuhring, T., Sieving, R. E., Shew, M., Ireland, M., Bearinger, L. H., & Udry, J. R. (1997). Protecting Adolescents from Harm. Findings from the National Longitudinal Study on Adolescent Health. *Journal of the American Medical Association*, 278(10), 823-832. <https://doi.org/10.1001/jama.1997.03550100049038>

Rogers, E. M., & Kincaid, D. L. (1980). *Communication Networks: Toward a New Paradigm for Research*. (1<sup>st</sup> ed.) The Free Press.

Schwimmer, E. (1973). *Exchange in the Social Structure of the Orokaiva: Traditional and Emergent Ideologies in the Northern District of Papua*. (1<sup>st</sup> ed.) C Hurst and Co-Publishers Ltd.

Scott, J. P. (1980). *The Anatomy of Scottish Capital: Scottish Companies and Scottish Capital*. (1<sup>st</sup> ed.) Croom Helm.

Smith, D. A., & White, D. R. (1992). Structure and Dynamics of the Global Economy: Network Analysis of International Trade 1965-1980. *Social Forces*, 70(4), 857-893. <https://doi.org/10.1093/sf/70.4.857>

Takahata, Y. (1991). *Diachronic Changes in the Dominance Relations of Adult Female Japanese Monkeys of the Arashiyama B Group*. (1<sup>st</sup> ed.) Albany: State University of New York Press.

Tsakas, N. (2016). On Decay Centgrality. *arXiv: 1604.05582*.

White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The Structure of the Nervous System of the Nematode *Caenorhabditis Elegans*. *Philosophical Transactions B*, 314(1165), 1-340. <https://doi.org/10.1098/rstb.1986.0056>

Zachary, W. W. (1977). An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research*, 33(4), 452-473.

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).