International Journal of Combinatorial Optimization Problems and Informatics, 11(3), Sep-Dec 2020, 76-96. ISSN: 2007-1558.

www.editada.org

Edge Similarity Index for Complex Network Analysis

Natarajan Meghanathan
Jackson State University, Jackson, MS 39217, USA.
E-mail: natarajan.meghanathan@jsums.edu

Abstract. We propose a novel network-level metric called Edge Similarity Index (ESI) to quantify the extent of similarity between any two edges of a complex network with respect to the values for a node-level metric (like centrality metric) of its end vertices. To assess the ESI measure for a complex real-world network with respect to a node-level metric, we propose to first construct a logical network whose vertices are the actual edges of the network (with coordinates corresponding to the normalized node-level metric values of the actual end vertices), and there exists a (logical) edge between two logical vertices if the Euclidean distance between their corresponding coordinates is within a threshold distance. We propose a binary search algorithm to determine the minimum value for this threshold distance ($dist_{thresh}^{\min}$) that would result in a connected logical unit-disk graph; the ESI value for the complex network is then computed as $1-\left(dist_{thresh}^{\min}/\sqrt{2}\right)$. The ESI values range from 0.0 to 1.0; the larger the ESI value with respect to a node-level metric, we claim that more similar are any two edges in the network with respect to the node-level metric values for their end vertices.
Keywords: Edge Similarity, Centrality Metrics, Threshold Distance, Binary Search Algorithm, Logical Graph

Article Info
Received Mar 13, 2019
Accepted Jan 2, 2020

1 Introduction

Similarity assessment is an important problem in complex network analysis. Until now, most of the focus has been on assessing the similarity among the nodes in the network at two different levels: between any two nodes (measures such as cosine similarity [1], matching index [2], etc) or between a group of nodes (measures such as equivalence classes [3]) in the network. Similarity assessment among the nodes is typically conducted on the basis of their values for the topological metrics (such as centrality metrics [3]) and/or the domain-level metrics (such as age, height, number of publications, etc). Throughout the paper, the terms 'vertex' and 'node', 'edge' and link', 'network' and 'graph' are used interchangeably. They mean the same.

The assortative index (abbreviated as ASI throughout this paper) [4] measure is the only prominent measure available in the literature for similarity assessment with regards to the edges. However, the assortative index measure just captures the extent to which the values for the end vertices of any edge are similar to each other with respect to a node-level metric. The ASI of a network with respect to a node-level metric is computed as the Pearson's correlation coefficient (ranges from -1 to 1) [5] of the node-level metric values of the end vertices of the edges in the network. With such a formulation for the ASI measure, we can only assess whether the node-level metric value for one end vertex (say, u) of an edge (u, v) would be similar or dissimilar to the node-level metric value for the other end vertex (say, v) of the edge (u, v). An assortative network is the one whose ASI value (with respect to a node-level metric) is positive, and the end vertices of the edges in an assortative network are considered to exhibit similar values with respect to the node-level metric. On the other hand, the end vertices of the edges in a dissortative network (with a negative ASI value) are considered to exhibit dissimilar values for the node-level metric. Networks whose ASI value is closer to 0 are neither assortative nor dissortative.

Our focus in this paper is on assessing the similarity between any two edges in the network. Given two edges (u1, v1) and (u2, v2), the ASI measure cannot quantify the extent to which the node-level metric values of the end vertices u1 and v1 would be similar to the node-level metric values of the end vertices u2 and v2. The ASI measure could only quantify how similar would the node-level metric value for u1 would be similar to that of v1 (and likewise the similarity/dissimilarity of the node-level metric values for u2 and v2). Consider a hub-and-spoke network below (Figure 1.a) wherein six-spoke vertices (each of degree one) are connected to a hub vertex (of degree six). The ASI for such a hub-and-spoke network will be -1.0; i.e., the network is highly dissortative with respect to node degree. On the other hand, any two edges in this hub-and-spoke network are exactly similar to each other on the basis of a tuple representing the degree values of the end vertices. More precisely, if we were to represent the edges of the network as vertices in a coordinate system (wherein the coordinates are the degrees of the end vertices of an edge), all the six edges in the hub-and-spoke network could be represented by a tuple (1, 6) and will appear co-located. Consider another example: a ring network of seven vertices as shown in Figure 1.b, wherein the end vertices of all the edges are of the same degree (i.e., similar to each other with respect to node degree) and hence the network is highly assortative (ASI value of 1.0). When represented in a coordinate system of the degrees of the end vertices, all the seven edges could be represented by a tuple (2, 2) and will also appear co-located.

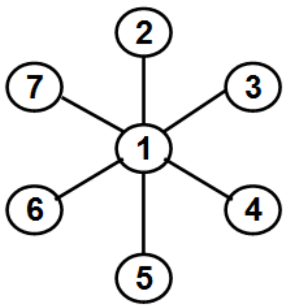


Figure 1.a: Hub-and-Spoke Network
ASI = -1.0 and ESI = 1.0

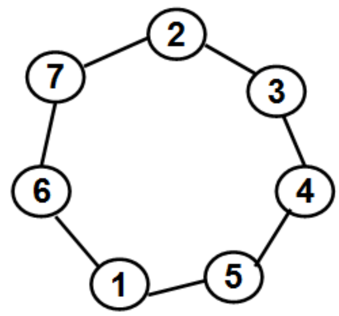


Figure 1.b: Ring Network
ASI = 1.0 and ESI = 1.0

Figure 1: Motivating Examples to Illustrate the Difference between Assortativity Index (ASI) and Edge Similarity Index (ESI)

The networks in Figures 1.a and 1.b represent the two extremes of assortativeness, but the edges in each of these networks would appear co-located when plotted with respect to a degree. Thus, ASI cannot capture the extent of similarity between any two edges of a network with respect to any node-level metric. We need a new "network-level" quantitative measure that could comprehensively capture the extent of similarity between any two edges of a network with respect to a node-level metric. We propose the name Edge Similarity Index (ESI) for such a measure. The ESI values (computed based on the procedure described below and in Section 2) for the networks in both Figures 1.a and 1.b with respect to node degree are 1.0 each. Hence, it is possible for both assortative and dissortative networks to have larger ESI values. In other words, ASI and ESI are independent of each other.

An outline of our approach to determine the ESI value for a network with respect to a particular node-level metric is as follows (more details are in Section 2): We first distribute the edges of the network as data points (vertices) in a logical two-dimensional coordinate system wherein the coordinates of a logical vertex are the normalized node-level metric values of the actual end vertices of the corresponding edge in the network. We seek to build a unit-disk graph of the logical vertices such that there exists an edge between two logical vertices if the Euclidean distance between them in the two-dimensional coordinate system is less than or equal to a threshold distance. The unit-disk graph would be completely connected if the threshold distance is $\sqrt{2}$ (where '2' corresponds to the number of dimensions in the coordinate system) and would not be connected if the threshold distance is 0 (unless all the vertices are co-located as in Figures 1.a and 1.b). We use a binary search approach to determine the minimum value for the threshold distance ($dist_{thresh}^{\min}$) that would yield a connected unit-disk graph of the logical vertices. The smaller the $dist_{thresh}^{\min}$ value (ranges from 0 to $\sqrt{2}$), the closer are the data points (logical vertices) to each other in the logical two-dimensional coordinate system (i.e., more similar are the edges in the actual network with respect to the node-level metric values for the end vertices) and vice-versa. Hence, we formulate the edge similarity index (ESI) metric $1-\left(dist_{thresh}^{\min}/\sqrt{2}\right)$. The ESI value (ranges from 0 to 1) for a network could be computed with respect to any node-level metric (say, topology-based centrality metrics or domain-level metrics). Also, the ESI values of a network with respect to two different node-level metrics need not be the same.

The proposed ESI metric can be useful for several applications; a sample list is as follows: (1) A larger ESI value for a communication network with respect to router capacities (node-level metric) is an indication that the logical vertices (links in the actual network) in the logical two-dimensional coordinate system are closer/similar to each other. Packets propagating through such similar links are likely to experience less jitter (i.e., less variation in the end-to-end packet delay) [6]. (2) A larger ESI value for a social network with respect to any domain-level node metric (such as age, height, salary, etc) is an indication that any two links of users are likely to be similar to each other (for example, majority of the links are between the younger users or majority of the links are between a taller user and a shorter user). (3) A larger ESI value for a collaboration network of researchers with respect to a domain-level node metric (such as the h-index of the researchers, the total dollar value of their grants, etc) is an indication that the researchers of any chosen link are very much comparable to the researchers of another link (for example, majority of the links could be between a researcher with a lower h-index and a researcher with a larger h-index, or majority of the links could be between researchers with a larger dollar value for their grants, etc). (4) One could run a clustering algorithm on the connected unit-disk graph of the logical vertices (actual edges) and identify smaller clusters of logical vertices, if any exist, which would correspond to edges that are different from the majority of the edges in the actual network. Such edges could be construed as outliers (for example, strange associations in a social network) and could be removed from the network to impart larger homophily [7] with respect to the similarity of the links. (5) If a complex network has a larger ESI value with respect to a node-level metric, we could use the nature of the values of the end vertices of the edges in the network as the basis for predicting a link between any two vertices that are not yet connected. For example, if the majority of the links in a social network are between two taller people, then we could predict links between two taller people who are not yet connected to each other.

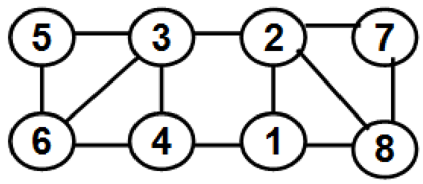
The rest of the paper is organized as follows: In Section 2, we propose the notion of Edge Similarity Index (ESI), explain its computation procedure (including the binary search algorithm) and analyze its time and space complexities. We also illustrate the computation of the ESI metric with an example. In Section 3, we review related work on similarity assessment of edges in complex networks. In Section 4, we first provide an overview of the 70 real-world networks that are used in the ESI analysis and present their ESI values with respect to four major centrality metrics (neighbourhood-based degree and eigenvector centrality metrics and the shortest path-based betweeness and closeness centrality metrics). We then assess the relationship between the ESI values of the real-world networks vs. their ASI values with respect to each of the above four centrality metrics as well as vs. the classical metrics that are a measure of the node density and variation in node degree. Section 5 concludes the paper and highlights its contributions to the literature.

2 Edge Similarity Index (ESI)

The Edge Similarity Index (ESI) is a network-level measure (with respect to a chosen node-level metric) of the similarity of the values for the end vertices between any two edges in the network. The ESI values of a network for two different node-level metrics need not be the same. In this paper, we use the four major centrality metrics (neighbourhood-based degree centrality: DEG and eigenvector centrality: EVC [6], and shortest path-based betweenness centrality: BWC [7-8] and closeness centrality: CLC [9-10]) as the node-level metrics on the basis of each of which we compute the ESI values. The graph in Figure 2 is used as a running example throughout this section to illustrate the procedure to compute the ESI value for a network with respect to a node-level metric.

2.1 Node-Level Centrality Metrics

Centrality metrics quantify the topological importance of a vertex in the network [3]. Centrality metrics could be broadly categorized as neighbourhood-based and shortest path-based. Though several centrality metrics have been proposed for the two categories, the degree and eigenvector centrality metrics are considered the representative (prototypical) metrics [11] for the neighbourhood-based category and the betweenness and closeness centrality metrics are considered the prototypical metrics for the shortest path-based category. The degree centrality (DEG) of a vertex is the number of neighbours of the vertex. The eigenvector centrality (EVC) of a vertex [6] is a measure of the degree of the vertex as well as the degrees of its neighbours. It is computed using the Power Iteration algorithm [6]. The betweenness centrality (BWC) of a vertex [7] is a measure of the fractions of the shortest paths (between any two vertices) that go through the vertex. The BWC values of the vertices are computed using the Brandes' algorithm [8]. The closeness centrality (CLC) of a vertex [9] is a measure of the distance (typically measured as the number of edges on the shortest path; determined using the Breadth-First Search algorithm [10]) of the vertex to the rest of the vertices in the network. In Figure 2, we show the raw values as well as the normalized values for the four centrality metrics (DEG, EVC, BWC and CLC) of the vertices. We normalize the raw node-level values for a metric by using the square root of the sum of the squares approach. That is, we first find the square root of the sum of the squares of the raw values of the vertices and divide each of the raw values for the vertices by the above square root value.



Raw Values of the Centrality Metrics

ID	DEG	EVC	BWC	CLC
1	3	0.3628	2.67	0.0833
2	4	0.4416	7.50	0.1000
3	4	0.4416	7.50	0.1000
4	3	0.3628	2.67	0.0833
5	2	0.2473	0.00	0.0667
6	3	0.3349	0.83	0.0769
7	2	0.2473	0.00	0.0667
8	3	0.3349	0.83	0.0769

Normalized Values of the Centrality Metrics

ID	DEG	EVC	BWC	CLC
1	0.3441	0.3628	0.2359	0.3565
2	0.4588	0.4416	0.6626	0.4280
3	0.4588	0.4416	0.6626	0.4280
4	0.3441	0.3628	0.2359	0.3565
5	0.2294	0.2473	0.0000	0.2854
6	0.3441	0.3349	0.0733	0.3291
7	0.2294	0.2473	0.0000	0.2854
8	0.3441	0.3349	0.0733	0.3291

Figure 2: Example Graph as well as the Raw and Normalized Values for the Centrality Metrics

2.2 Logical Topology of the Edges

We build a logical topology of the edges by distributing them as data points (vertices) in a two-dimensional coordinate system of the normalized node-level metric values (ranging from 0 to 1) of the end vertices of the edges. We represent an edge as the tuple (u, v), wherein u and v are the end vertices of the edge; the two dimensions in the coordinate system are referred to as the U-dimension and the V-dimension in Figure 3. There exists an edge (referred to as a logical edge) between two logical vertices in the two-dimensional coordinate system if the Euclidean distance between the coordinates of the two logical vertices is within a threshold distance, which could range from 0 to $\sqrt{2}$ (as the values for each of the two coordinate systems could range from 0 to 1). Figure 3 presents the distribution of the edges (as data points) from the example graph of Figure 1 on the basis of the normalized centrality values for their end vertices with respect to each of DEG, EVC, BWC and CLC. In each case, the two dimensions are the normalized centrality values of the end vertices of the edges. We observe the data points with respect to the CLC metric to be the closest to each other and those with respect to BWC to be the farthest from each other. This implies, the edges of the example graph of Figure 1 are relatively more similar to each other on the basis of the CLC values of the end vertices and least similar to each other on the basis of the BWC values of the end vertices. Though both BWC and CLC are shortest path-based centrality metrics, we observe such a trend for the real-world networks too (see Section 4).

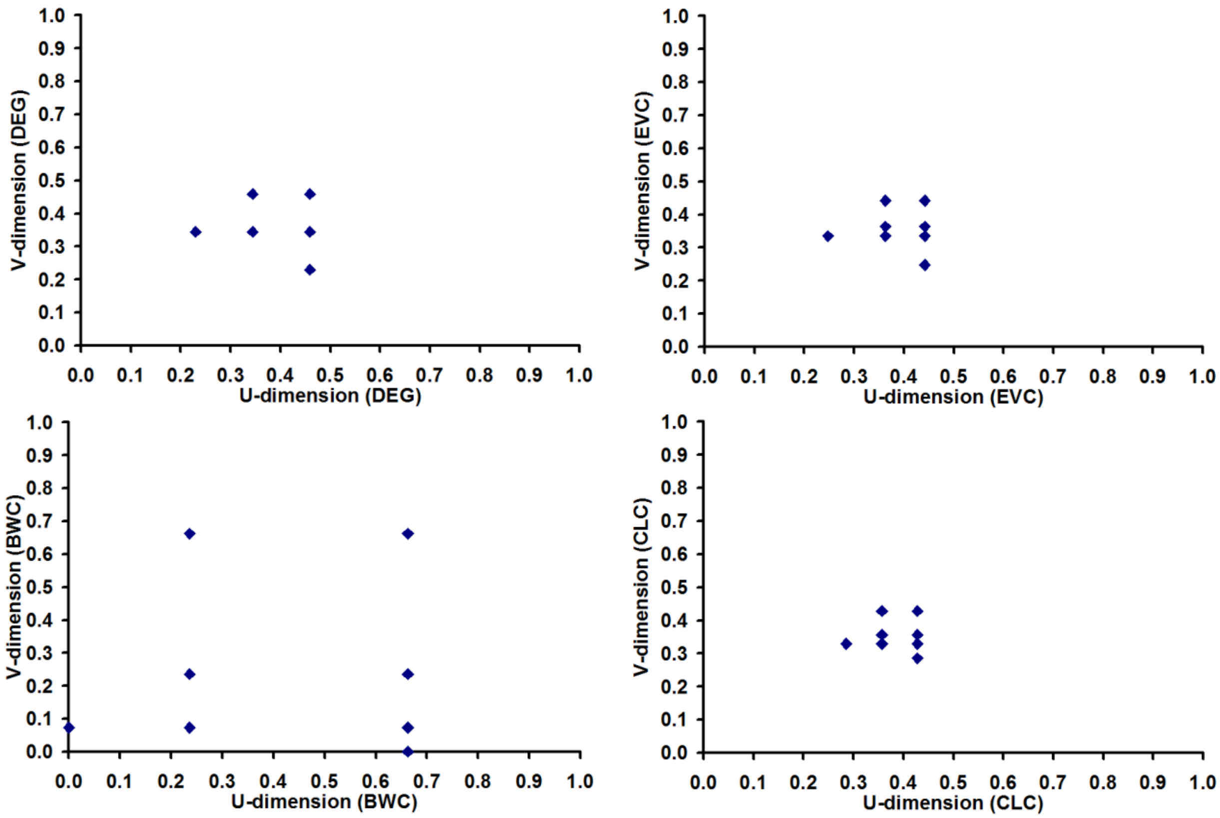


Figure 3: Logical Topologies of the Edges with respect to the Normalized Values for the Centrality Metrics of the End Vertices

2.3 Binary Search Algorithm

We seek to build a connected unit-disk graph of the logical vertices (edges in the actual network) in the two-dimensional coordinate system of subsection 2.2. Two logical vertices in the coordinate system are said to be connected through a logical edge if the Euclidean distance between them is within a threshold distance. The range for such a threshold distance in the two-dimensional coordinate system of the normalized centrality values is [0, ..., $\sqrt{2}$]. Of course, the logical vertices will be connected to each other if the threshold distance is $\sqrt{2}$. Unless the logical vertices are co-located at the same coordinate values, the minimum threshold distance for which we can get a connected unit-disk graph of the logical vertices will be greater than 0. Hence, we need to determine a minimum value for the threshold distance ($dist_{thresh}^{\min}$) that would contribute to a connected unit-disk graph of the logical vertices. We propose to use a binary search algorithm for this purpose (see Figure 4 for the pseudo-code).

Inputs

Logical topology L of the edges (distributed as vertices) corresponding to a real-world network G^R
Cutoff parameter ε
// Let C be the node-level metric used
// Let the coordinates of a vertex in L (say, edge (u, v) in G^R) with respect to C be represented as (u^C, v^C)

Output

Minimum threshold distance, $dist_{thresh}^{\min,C}$, with respect to the node-level metric C

Auxiliary Variables

Left Index (initialized to 0); Right Index (initialized to $\sqrt{2}$);Middle Index

Begin Binary Search Algorithm

while (| Right Index - Left Index | > ε) **do**
Middle Index = (Left Index + Right Index) / 2
Build a unit-disk graph G^L of the vertices of L using the Middle Index as the threshold distance
/* Two logical vertices ($u1$, $v1$) and ($u2$, $v2$) are connected in G^L if the Euclidean distance
$\sqrt{(u_1^C - u_2^C)^2 + (v_1^C - v_2^C)^2} \leq$ Middle Index */

if (G^L is connected) **then**
Right Index = Middle Index
else
Left Index = Middle Index
end if

end while

return $dist_{thresh}^{\min,C}$ = Right Index

End Binary Search Algorithm

Figure 4: Binary Search Algorithm to Find the Minimum Threshold Distance for a Connected Logical Unit-Disk Graph with respect to the Normalized Values for a Node-Level Metric

The binary search algorithm goes through a sequence of iterations until it determines $dist_{thresh}^{\min}$. In all the iterations, the algorithm maintains a left index (initialized to 0) and right index (initialized to $\sqrt{2}$) such that the unit-disk graph of the logical vertices is not connected when the left index is used as the threshold distance and is connected when the right index is used as the threshold distance. In each iteration, we determine a middle index = (left index + right index) / 2, and we build a unit-disk graph of the logical vertices such that two logical vertices are connected with an edge if the Euclidean distance between them is less than or equal to the middle index. If the unit-disk graph corresponding to the middle index as the threshold distance is connected, we set right index = middle index; otherwise, we set left index = middle index and continue to the next iteration. By doing so, we not only maintain the invariant (mentioned above), we also reduce the search space by half in each iteration. Finally, when the difference between the right index and left index is less than or equal to a cutoff parameter (ε), we stop the algorithm and declare the value of the right index at the time of exiting the while loop as the value for the parameter $dist_{thresh}^{\min}$. The ESI of the real-world network with respect to the node-level metric is then computed as $1-\left(dist_{thresh}^{\min} / \sqrt{2}\right)$. Note that we set $dist_{thresh}^{\min}$ to be the value of the right index (rather than the middle index) at the time of exiting the loop, as the unit-disk graph is guaranteed to be connected when the right index value is used as the threshold distance (and the graph need not be connected when the middle index value is used as the threshold distance).

The number of iterations of the proposed binary search algorithm simply depends on the number of dimensions used in the coordinate system (which is two) and the value for the cutoff parameter, ε. To begin with, the minimum threshold distance can be anywhere in the range (0, ..., $\sqrt{2}$], and the left index, right index and middle index are 0, $\sqrt{2}$ and $\sqrt{2}$/2 = 0.7070 respectively. At the end of the first iteration, we reduce the search space from (0, ..., $\sqrt{2}$] to either (0, ..., 0.7070] or (0.7070, ..., $\sqrt{2}$]. At the end of the second iteration, we reduce the search space to one of these ranges: (0, ..., 0.3535], (0.3535, ..., 0.7070], (0.7070, ..., 1.0605] or (1.0605, ..., $\sqrt{2}$]. We continue the process until the difference between the right index and left index is less than or equal to ε. As we reduce the search space by half in each iteration, the number of iterations it would take for a search space size of $\sqrt{2}$ to reduce to a search space size of ε can be simply given by $\log_2^{\sqrt{2}/\epsilon}$. This implies that the number of iterations needed for the binary search algorithm to determine $dist_{thresh}^{\min}$ does not depend on the number of nodes and edges in the real-world network graph considered or on the node-level metric used as the basis for the values of the end vertices of the edges.

The time complexity for each iteration of the algorithm depends on the time complexity to build the logical graph (GL) and check for its connectivity. Note that the number of logical vertices in GL corresponds to the actual number of edges (say, denoted E) in the real-world network GR. It takes O(E2) time to check for edges between any two logical vertices in GL and a time complexity O(E) to run the Breadth-First Search algorithm to check for the connectivity of GL. The time complexity of an iteration of the binary search algorithm is thus dominated by the time complexity to build a logical graph of the E vertices.

Hence, the overall time complexity of the binary search algorithm is O(E2* $\log_2^{\sqrt{2}/\epsilon}$).

The space complexity of the binary search algorithm depends on the memory required to store the logical graph built in an iteration. Note that the logical graph built for an iteration could be cleared at the end of the iteration. As the logical graph would be of E vertices and E2 edges (at the worst-case), the space complexity of the algorithm is O(E2), where E is the number of edges in the real-world network graph.

2.4 Example to Illustrate the Execution of the Binary Search Algorithm

In this subsection, we illustrate the execution of the binary search algorithm on the example graph of Figure 1 with respect to the degree centrality metric. Figure 5 presents a sequence of logical graphs that represent the unit-disk graph of the logical topology of vertices (edges in the actual graph) for different values of the middle index (threshold distance, abbreviated as TD in Figure 5) encountered during the binary search algorithm. The IDs of the vertices in the logical graphs are represented as tuples corresponding to the end vertices of the edges in the actual graph. We start with a left index of 0 and right index of 1.41421 (~ $\sqrt{2}$), the values are rounded to the fifth decimal precision. The middle index for the first iteration is 0.70710, the average of 0 and 1.41421. As the unit-disk graph is connected for a threshold distance of 0.70710, we continue by discarding the right search space and set the middle index value of 0.70710 to be the new value of the right index. The second iteration is then run based on a middle index of (0 + 0.70710) / 2 = 0.35355 for which the unit-disk graph is connected. We move on to the third iteration by setting the latest middle index (0.35355) to be the new value of the right index. We continue like this until the difference between the right index and left index is less than or equal to the cutoff parameter ε , which is 0.01 for all the analysis conducted and presented in this paper.

Table 1 presents the values for the left index, right index and middle index for each iteration, and also reports whether the logical unit-disk graph is connected or not for the middle index (threshold distance) used in an iteration. During the beginning of the 9th iteration, we find the left index to be 0.11048 and the right index to be 0.11600, and their difference is less than 0.01. Hence, we stop the algorithm and conclude the value of $dist_{thresh}^{\min,DEG}$ to be the latest value of the right index, which is 0.11600. The ESI of the example graph with respect to the DEG centrality metric is then computed as $1-\left(dist_{thresh}^{\min,DEG} / \sqrt{2}\right)$ = 1− (0.11600 / $\sqrt{2}$) = 0.91797.

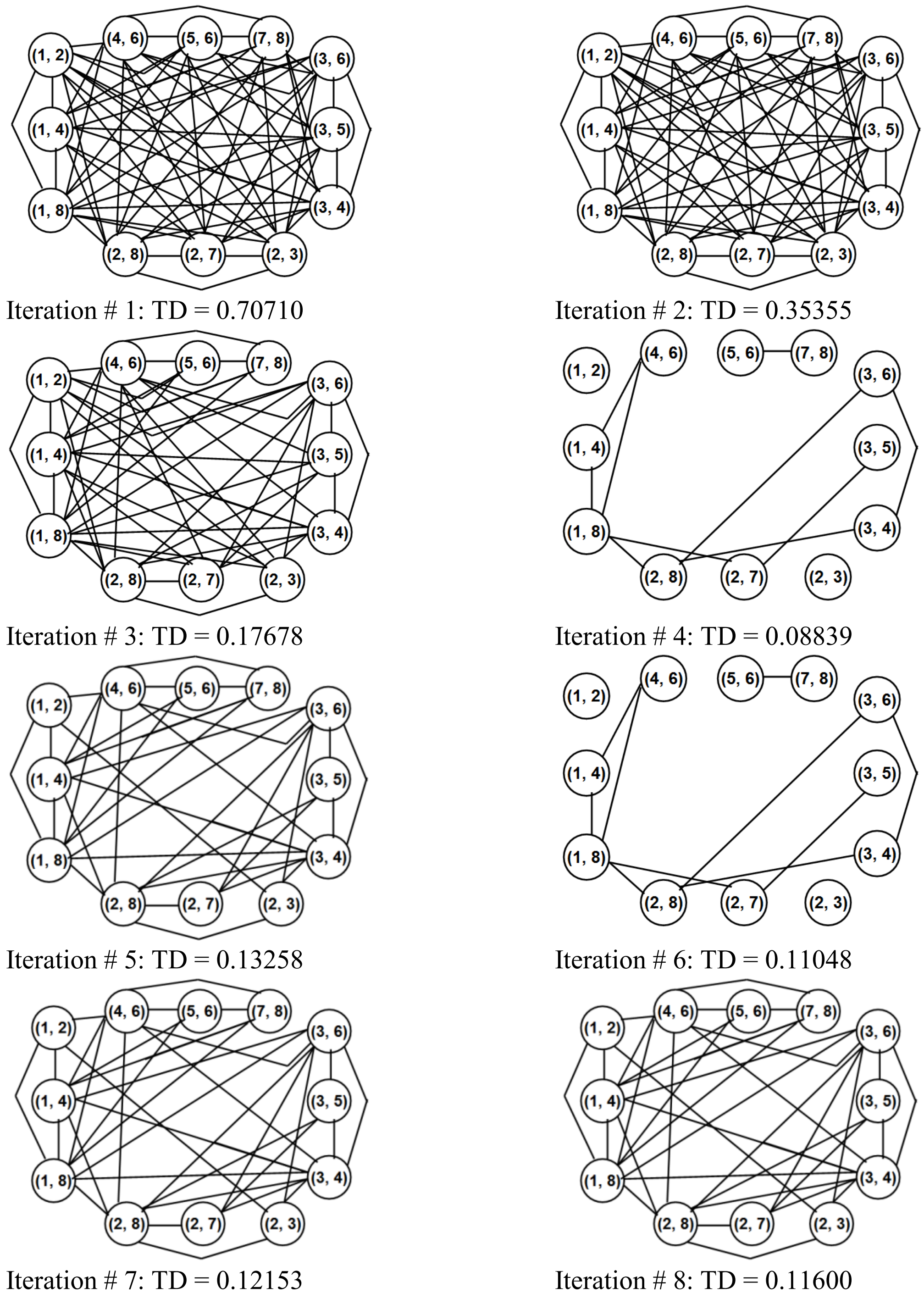


Figure 5: Sequence of Logical Unit-Disk Graphs Resulting from the Execution of the Binary Search Algorithm on the Example Graph of Figure 1, with respect to the Degree Centrality Metric

By conducting a similar execution of the algorithm with respect to the EVC, BWC and CLC metrics, we observe the corresponding ESI Values to be 0.93359, 0.69531 and 0.94922 respectively (see Figure 6). These ESI values clearly depict our earlier observation in subsection 2.2 that the edges of the example graph are most similar to each other with respect to the CLC metric and the most dissimilar to each other with respect to the BWC metric. We could observe that a smaller magnitude of difference in the ESI values itself manifests to an appreciable extent of difference in the similarity of the edges with respect to their distribution in the logical topology. In Figure 6, we also show the Assortativity Index (ASI) values of the edges with respect to each of the four centrality metrics. We observe the example graph of Figure 1 to be relatively more dissortative with respect to DEG (ASI = −0.19365) and neutral with respect to BWC (ASI = 0.03052) and CLC (ASI = −0.04239) . However, the ESI values for CLC and BWC are widely different from each other.

Table 1: Iteration Details of the Execution of the Binary Search Algorithm on the Example Graph of Figure 1, with respect to the Degree Centrality Metric

It #	Left Index	Right Index	Right Index - Left Index	Middle Index (Threshold Distance)	Connectivity of the Logical Unit Disk Graph
1	0	1.41421	1.41421	0.70710	Connected
2	0	0.70710	0.70710	0.35355	Connected
3	0	0.35355	0.35355	0.17678	Connected
4	0	0.17678	0.17678	0.08839	Not connected
5	0.08839	0.17678	0.08839	0.13258	Connected
6	0.08839	0.13258	0.04419	0.11048	Not connected
7	0.11048	0.13258	0.02210	0.12153	Connected
8	0.11048	0.12153	0.01105	0.11600	Connected
9	0.11048	0.11600	0.00552 < 0.01 STOP!!!		

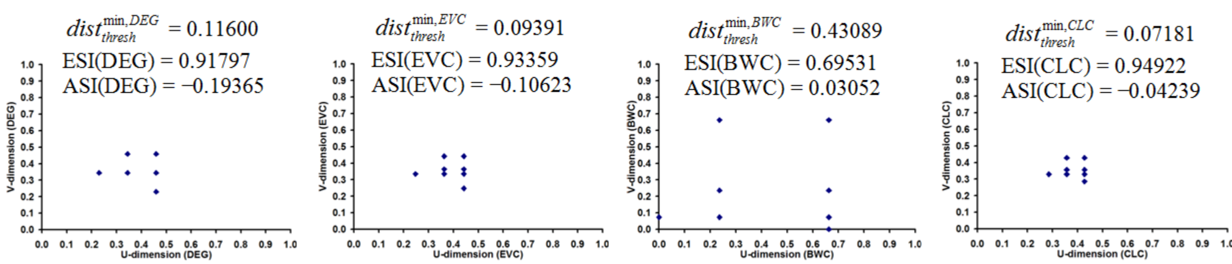


Figure 6: Comparison of the ESI and ASI Values for the Example Graph of Figure 1 with respect to the Four Centrality Metrics: DEG, EVC, BWC and CLC

3 Related Work and Our Contributions

In this section, we review the related work that has been proposed in the literature for quantifying edge similarity and/or node similarity in complex networks. The Assortativity Index (ASI) [4] is the most prominent measure proposed in the literature to quantify the similarity of the end vertices of the edges in a complex network with respect to a node-level metric. If the end vertices of the edges have similar values for the node-level metric, then the network is considered to be assortative (i.e., similar vertices are connected to similar vertices) with respect to the node-level metric and the ASI of the network with respect to the metric is closer to 1.0. On the other hand, if the end vertices of the edges have dissimilar values for the node-level metric, the network is said to be dissortative with respect to the metric and the ASI of the network will be closer to -1.0. More than 60% of the 50 real-world networks analyzed in [12] were observed to be neither assortative nor dissortative with respect to the degree centrality metric. In general, real-world networks are more likely to be neutral with respect to the DEG and BWC metrics, and assortative with respect to the EVC and CLC metrics [12]. On the other hand, in the results reported in Section 4, we observe several real-world networks that are neutral with respect to assortativity to incur larger ESI(DEG) and ESI(BWC) values.

Several related works on assortativity are also available in the literature. The local assortativity measure [13] quantifies the contribution of a node to the assortativity of the network. With the local assortativity measure, we could identify nodes that are assortative (i.e., connected to similar nodes) in a dissortative network and vice-versa. In [2], the maximal assortative and maximal dissortative matching of the edges in complex real-world networks were studied with respect to node degree. Edges with similar degree values for the end vertices were observed to be preferable for maximal assortative matching, whereas edges with dissimilar degree values for the end vertices were observed to be preferable for maximal dissortative matching. In addition, the algorithms for maximal assortative or maximal dissortative matching gave secondary preference to edges with fewer adjacent edges so that the number of edges included in the matching could also be maximized.

All the above schemes related to assortativity are focused on identifying edges with similar (or dissimilar) values for the end vertices, on a per-edge basis or on a per-neighbourhood basis. We do not expect any coincidence between the edges that are part of a maximal assortative or maximal dissortative matching with those edges that are part of the connected logical unit-disk graph for a minimum value of the threshold distance. For example, any two edges that are part of a maximal assortative matching could still have dissimilar values for their end vertices (one edge could connect two high degree vertices and another could connect two low degree vertices). On the other hand, we could run a clustering algorithm on the connected logical unit-disk graph corresponding to the minimum threshold distance and obtain clusters of similar edges with respect to the values for the end vertices; we could pick non-adjacent edges from each such cluster and use these edges as part of a maximal assortative matching with similar values for the end vertices.

Note that measures such as edge degree (sum of the degrees of the end vertices minus 2) [14], edge Jaccard Index (ratio of the number of common neighbours and the total number of neighbours excluding the two end vertices) [15], etc cannot be used to assess the similarity between edges as such measures cannot capture the relative proximity of the two edges in a coordinate system of the degrees of their end vertices. For example, two edges with degree tuples (2, 10) and (10, 2) would have the same edge degree (10), but would be located far away from each other in the coordinate system of the degrees of the end vertices of the edges. Two edges with identical Jaccard Index ratio of 2/3 = 0.67 = 4/6 could actually comprise of end vertices with different degrees that are not co-located in the coordinate system. To the best of our knowledge, ours is the first work to use a coordinate system for the node-level metric values of the end vertices to assess the similarity of the edges in a complex network.

4 Real-World Networks

Table 2 presents the 70 real-world networks (each network is identified with a unique three-character code) that are analyzed in this paper for the proposed edge similarity index (ESI) measure. The real-world networks are spread over a total of 13 different domains; the domain names and the number of networks for each of these domains are as follows: Biological network (15), Acquaintance network (12), Friendship network (11), Co-appearance network (8), Employment network (6), Citation network (4), Literature network (3), Transportation network (3), Collaboration network (2), Game network (2), Political network (2), Geographical network (1) and Trade network (1). A brief description of the typical nature of the nodes and edges in the major domains (for which at least 6 real-world networks are listed in Table 2) is as follows: In a biological network, the nodes could be the genes, proteins and their associated transcriptions and the edges could be the interactions between these nodes. A biological network could also model organisms (nodes) of a particular species and their interactions (edges). An acquaintance network comprises of individuals (nodes) who slightly know each other and their interaction (edge) is captured during an observation period. A friendship network comprises of individuals (nodes) who know each other very well and there is no need for an observation period to capture their friendship (edges). An employment network comprises of individuals (nodes) who interact (edges) at the official level due to their job requirements and according to the policies of the organization, they are working for and not at the personal level. A co-appearance network is typically a network of novel characters or dictionary words (nodes) who co-appear alongside each other (modelled as edges).

In Table 2, we present the number of nodes and edges for each real-world network along with two classical parameters: the edge density (ρedge) and the spectral radius ratio for node degree (λsp) [16]. The edge density for a network (ranging from 0...1) is computed as the ratio of the actual number of edges in the network and the maximum possible number of edges in the network. For a network of N nodes, the maximum possible number of edges in the network is N(N-1)/2. The spectral radius ratio for node degree (λsp ≥ 1) for a network [16] is a measure of the extent of variation in node degree and is computed as the ratio of the spectral radius of the adjacency matrix of the network and the average node degree. The larger the λsp value for a network, the larger is its variation in node degree. Real-world networks whose degree distribution models the power-law (scale-free networks) [17] incur a larger λsp value, whereas real-world networks whose degree distribution models a Poisson distribution (random networks) [18] incur a lower λsp value.

Table 2: Real-World Networks studied for Edge Similarity Analysis

Seq	Net.	Net. Description	Ref.	Network Domain	λ_{sp}	#nodes	#edges	ρ_{edge}
1	MDN	Macaque Dominance Net.	[19]	Biological Network	1.04	62	1167	0.6171
2	CAT	Cat Brain Network	[20]	Biological Network	1.19	65	730	0.3510
3	HCG	Hepatitus C Genetic Int. Net.	[21]	Biological Network	4.17	105	123	0.0225
4	FFW	Florida Food Web Net.	[22]	Biological Network	1.22	128	2106	0.2591
5	FBF	Flensburg Food Web Net.	[23]	Biological Network	1.80	180	1577	0.0979

6	HHG	Human Herpes 4 Genet Net	[21]	Biological Network	6.07	216	260	0.0112
7	CEN	C. Elegans Neural Network	[24]	Biological Network	1.68	297	2148	0.0489
8	GGI	Gallus Genetic Interact. Net.	[25]	Biological Network	7.00	313	364	0.0075
9	CEM	Celegans Metabolic Net.	[26]	Biological Network	2.94	453	2025	0.0198
10	XGI	Xenopus Genetic Inter. Net.	[25]	Biological Network	7.53	461	578	0.0055
11	RTN	Rat Transcription Network	[27]	Biological Network	3.40	488	1092	0.0092
12	TTN	MTuberculosis Trans. Net.	[27]	Biological Network	6.13	756	937	0.0033
13	YTH	Yeast Two-Hybrid PPI Net.	[28]	Biological Network	4.29	813	843	0.0026
14	HIV	Human HIV Gen. Inter. Net.	[29]	Biological Network	6.16	1005	1189	0.0024
15	MTN	Mouse Transcription Net.	[30]	Biological Network	4.30	1130	2403	0.0038
16	TEN	Taro Exchange Network	[31]	Acquaintance Network	1.06	22	39	0.1688
17	SSM	Sawmill Strike Comm. Net.	[32]	Acquaintance Network	1.22	24	38	0.1377
18	KCN	Karate Club Network	[33]	Acquaintance Network	1.47	34	78	0.1390
19	KFP	Korea Family Planning Net.	[34]	Acquaintance Network	1.70	37	85	0.1412
20	CDF	College Dorm Fraternity Net	[35]	Acquaintance Network	1.11	58	967	0.5850
21	DON	Dolphin Network	[36]	Acquaintance Network	1.40	62	159	0.0841
22	MTB	Madrid Train Bombing Net.	[37]	Acquaintance Network	1.95	64	295	0.2012
23	SJN	San Juan Sur Family Net.	[38]	Acquaintance Network	1.29	75	155	0.0559
24	HTN	Hypertext 2009 Network	[39]	Acquaintance Network	1.21	115	2164	0.3418
25	PSN	Primary School Contact Net.	[40]	Acquaintance Network	1.22	238	5539	0.1964
26	DRN	Drug Network	[41]	Acquaintance Network	2.76	212	284	0.0076
27	ISP	Infectious Socio-Patterns Net	[39]	Acquaintance Network	1.69	309	1924	0.0404
28	MMN	ModMath Network	[42]	Friendship Network	1.59	30	61	0.1025
29	FHT	Friendship in Hi-Tech Firm	[43]	Friendship Network	1.57	33	91	0.2333
30	WSB	Windsurfers Beach Network	[44]	Friendship Network	1.22	43	336	0.3721
31	TWF	Teenage Female Friend Net.	[45]	Friendship Network	1.49	47	77	0.0996
32	PFN	Prison Friendship Network	[46]	Friendship Network	1.32	67	142	0.0823
33	UKF	UK Faculty Friendship Net.	[47]	Friendship Network	1.35	83	578	0.1781
34	AFB	Author Facebook Network	-	Friendship Network	2.29	171	940	0.0661
35	FMH	Faux Mesa High School Net	[48]	Friendship Network	2.81	147	202	0.0193
36	RHF	Residence Hall Friend Net.	[49]	Friendship Network	1.27	217	1839	0.0785
37	CKM	CKM Physicians Network	[50]	Friendship Network	4.74	246	668	0.0222
38	FB2	Facebook Network 2	[51]	Friendship Network	13.69	324	2218	0.0424
39	HCN	Huckleberry Coappear. Net.	[52]	Co-appearance Network	1.66	76	302	0.1114
40	LMN	Les Miserables Network	[52]	Co-appearance Network	1.82	77	254	0.0868
41	CFN	Copperfield Network	[52]	Co-appearance Network	1.83	89	407	0.1085
42	ADJ	Word Adjacency Network	[53]	Co-appearance Network	1.73	112	425	0.0684
43	SMN	Slovenian Magazine Net.	[54]	Co-appearance Network	1.05	124	5972	0.7831
44	AKN	Anna Karnenina Network	[52]	Co-appearance Network	2.48	140	494	0.0522
45	MUN	Marvel Universe Network	[55]	Co-appearance Network	2.54	167	301	0.0222
46	ROG	Roget Network	[52]	Co-appearance Network	1.68	1022	3648	0.0070
47	FTC	Flying Teams Cade Net.	[56]	Employment Network	1.21	48	170	0.1507
48	CSA	CS Department Aarhus Net.	[57]	Employment Network	2.12	61	219	0.1197
49	LLF	Lazega Law Firm Net.	[58]	Employment Network	2.63	71	205	0.0825
50	MCE	Manufact. Comp. Empl. Net.	[59]	Employment Network	1.12	77	1549	0.7949
51	JBN	Jazz Band Network	[60]	Employment Network	1.45	198	2742	0.1406
52	SDI	Scotland Corp. Interlock Net	[61]	Employment Network	1.94	230	359	0.0121
53	CLN	Centrality Literature Net.	[62]	Citation Network	2.03	118	613	0.0742
54	GD96	Graph Drawing 1996 Net	[63]	Citation Network	2.38	180	228	0.0142
55	CGD	Citation Graph Drawing Net	[64]	Citation Network	2.24	259	640	0.0133
56	PDN	Perl Developers Network	[65]	Citation Network	5.22	839	2111	0.0060
57	DLN	Dutch Literature 1976 Net.	[66]	Literature Network	1.49	37	81	0.1345
58	GLN	Graph Glossary Network	[63]	Literature Network	2.01	67	118	0.0923
59	PBN	US Politics Books Network	[67]	Literature Network	1.42	105	441	0.0808

60	APN	US Airports 1997 Network	[63]	Transportation Network	3.22	332	2126	0.0387
61	LTN	London Transportation Net.	[68]	Transportation Network	3.60	381	507	0.0070
62	EUA	EU Air Transportation Net.	[69]	Transportation Network	3.81	418	1999	0.0229
63	ERD	Erdos Collaboration Net.	[63]	Collaboration Network	3.00	433	1314	0.0119
64	MSJ	Soc. Net. Journal Co-authors	[70]	Collaboration Network	3.48	475	625	0.0056
65	SWC	Soccer World Cup 1998 Net	[63]	Game Network	1.45	35	118	0.1983
66	FON	US Football Network	[71]	Game Network	1.01	115	613	0.0935
67	MPN	Mexican Political Elite Net.	[72]	Political Network	1.23	35	117	0.1966
68	SPR	Senator Press Release Net.	[73]	Political Network	1.57	92	477	0.1140
69	USS	US States Network	[74]	Geographical Network	1.25	49	107	0.0910
70	WTN	World Trade Metal Network	[75]	Trade Network	1.38	80	875	0.2769

Table 3 presents the ESI (Edge Similarity Index) and ASI (Assortativity Index) values of the real-world networks with respect to each of the four centrality metrics: DEG (degree), EVC (eigenvector), BWC (betweenness) and CLC (closeness). We notice the ESI(CLC) values to be significantly larger than the ESI values incurred with the other three centrality metrics. The ESI(CLC) values for all the 70 real-world networks are greater than 0.90. On the other hand, the ESI(BWC) values are the lowest for a majority of the real-world networks (59 of the 70 networks). Hence, we could say, the ESI values with respect to the BWC and CLC metrics are respectively the lower bound and upper bound for the ESI values incurred with any these four prototypical centrality metrics with a probability of 0.84 (= 59/70) and 1.00 respectively. The median of the ESI values with respect to DEG, EVC, BWC and CLC are 0.939, 0.943, 0.844 and 0.984 respectively. In Table 3, we highlight the cells (in yellow) for which the ESI value of a centrality metric is the lowest for a real-world network.

The ASI of a network/graph with respect to a node-level metric is calculated as the Pearson's correlation coefficient between two sets representing the metric values for the end vertices of the edges in the graph. For details of calculating the ASI of a graph, the reader is referred to [5]. Per [12], on the basis of the ASI values, the assortativeness of the real-world networks could be classified into three regimes: Dissortative regime ($-1 \leq ASI < -0.2$); Neutral regime ($-0.2 \leq ASI \leq 0.2$); and Assortative regime ($0.2 < ASI \leq 1.0$). The median of the ASI values reported in Table 3 with respect to DEG, EVC, BWC and CLC are -0.031, 0.232, -0.066 and 0.239. Overall (refer to Figure 8 for a visual presentation), we observe the real-world networks to be relatively more assortative with respect to EVC and CLC, and neutral (neither assortative nor dissortative) with respect to DEG and BWC.

Table 3: Edge Similarity Index (ESI) and Assortativity Index (ASI) of the Real-World Networks

Seq	Net.	Network Domain	Edge Similarity Index (ESI)				Assortativity Index (ASI)			
			DEG	EVC	BWC	CLC	DEG	EVC	BWC	CLC
1	MDN	Biological Network	0.980	0.980	0.965	0.992	-0.048	-0.016	-0.097	-0.047
2	CAT	Biological Network	0.965	0.973	0.891	0.984	0.003	0.105	-0.083	0.006
3	HCG	Biological Network	0.555	0.754	0.563	0.969	-0.382	-0.268	-0.351	-0.067
4	FFW	Biological Network	0.961	0.961	0.816	0.984	-0.119	0.014	-0.128	-0.096
5	FBF	Biological Network	0.914	0.961	0.645	0.984	-0.376	-0.085	-0.146	0.009
6	HHG	Biological Network	0.457	0.641	0.387	0.961	-0.214	-0.072	-0.123	0.306
7	CEN	Biological Network	0.879	0.973	0.477	0.992	-0.167	0.095	-0.137	0.162
8	GGI	Biological Network	0.574	0.664	0.582	0.977	-0.205	0.135	-0.190	0.360
9	CEM	Biological Network	0.797	0.910	0.512	0.988	-0.228	-0.196	-0.127	-0.003
10	XGI	Biological Network	0.578	0.734	0.566	0.984	-0.074	0.031	-0.072	0.182
11	RTN	Biological Network	0.883	0.891	0.781	0.973	-0.081	0.211	-0.081	0.502
12	TTN	Biological Network	0.609	0.566	0.633	0.980	-0.009	0.571	0.056	0.897
13	YTH	Biological Network	0.855	0.672	0.637	0.980	-0.039	0.527	0.045	0.964
14	HIV	Biological Network	0.395	0.570	0.379	0.984	-0.143	-0.044	-0.134	0.196
15	MTN	Biological Network	0.898	0.902	0.871	0.977	-0.091	0.223	-0.073	0.376
16	TEN	Acquaintance Network	0.957	0.953	0.891	0.984	-0.362	0.263	-0.162	0.231
17	SSM	Acquaintance Network	0.914	0.902	0.648	0.949	-0.022	0.499	0.038	0.323
18	KCN	Acquaintance Network	0.875	0.945	0.648	0.973	-0.477	-0.242	-0.358	-0.081
19	KFP	Acquaintance Network	0.914	0.938	0.918	0.973	0.241	0.534	0.171	0.547

20	CDF	Acquaintance Network	0.977	0.980	0.953	0.992	-0.115	-0.099	-0.103	-0.119
21	DON	Acquaintance Network	0.977	0.953	0.844	0.984	-0.044	0.643	0.123	0.527
22	MTB	Acquaintance Network	0.953	0.945	0.941	0.984	0.029	0.390	-0.126	0.350
23	SJN	Acquaintance Network	0.941	0.887	0.723	0.980	0.030	0.652	0.168	0.512
24	HTN	Acquaintance Network	0.973	0.984	0.754	0.988	-0.121	-0.100	-0.098	-0.121
25	PSN	Acquaintance Network	0.992	0.992	0.961	0.996	0.218	0.290	0.099	0.261
26	DRN	Acquaintance Network	0.938	0.887	0.895	0.957	0.111	0.578	0.216	0.920
27	ISP	Acquaintance Network	0.977	0.984	0.855	0.996	0.286	0.559	0.133	0.765
28	MMN	Friendship Network	0.852	0.875	0.848	0.980	0.276	0.639	0.102	0.451
29	FHT	Friendship Network	0.941	0.949	0.883	0.977	-0.069	0.241	-0.166	0.092
30	WSB	Friendship Network	0.980	0.980	0.895	0.992	-0.108	-0.039	-0.099	-0.066
31	TWF	Friendship Network	0.969	0.906	0.875	0.984	0.363	0.901	0.390	0.814
32	PFN	Friendship Network	0.961	0.949	0.875	0.984	0.160	0.545	0.081	0.462
33	UKF	Friendship Network	0.961	0.965	0.816	0.988	0.039	0.252	-0.037	0.152
34	AFB	Friendship Network	0.984	0.961	0.852	0.965	0.349	0.894	0.094	0.812
35	FMH	Friendship Network	0.945	0.941	0.891	0.949	0.120	0.670	0.164	0.933
36	RHF	Friendship Network	0.977	0.973	0.914	0.992	0.097	0.361	-0.001	0.240
37	CKM	Friendship Network	0.617	0.742	0.656	0.961	0.037	0.011	0.046	0.037
38	FB2	Friendship Network	0.977	0.953	0.727	0.996	0.257	0.570	0.030	0.467
39	HCN	Co-appearance Network	0.789	0.883	0.469	0.914	0.030	0.183	0.013	0.435
40	LMN	Co-appearance Network	0.871	0.961	0.563	0.977	-0.077	0.432	-0.024	0.213
41	CFN	Co-appearance Network	0.746	0.895	0.383	0.949	-0.166	-0.089	-0.070	-0.086
42	ADJ	Co-appearance Network	0.895	0.898	0.762	0.988	-0.097	0.035	-0.097	0.129
43	SMN	Co-appearance Network	0.977	0.979	0.980	0.992	-0.228	-0.221	-0.197	-0.229
44	AKN	Co-appearance Network	0.883	0.961	0.773	0.984	-0.081	0.094	-0.083	0.117
45	MUN	Co-appearance Network	0.914	0.871	0.715	0.953	-0.018	0.566	-0.013	0.881
46	ROG	Co-appearance Network	0.992	0.988	0.980	0.977	0.182	0.482	0.102	0.614
47	FTC	Employment Network	0.949	0.957	0.895	0.984	-0.014	0.463	-0.052	0.238
48	CSA	Employment Network	0.855	0.914	0.766	0.973	-0.115	-0.154	-0.091	-0.107
49	LLF	Employment Network	0.625	0.762	0.613	0.926	-0.004	-0.011	-0.004	-0.004
50	MCE	Employment Network	0.969	0.980	0.844	0.980	-0.040	0.092	-0.072	-0.063
51	JBN	Employment Network	0.965	0.980	0.727	0.992	0.031	0.353	-0.038	0.180
52	SDI	Employment Network	0.961	0.785	0.902	0.961	0.083	0.950	0.221	0.882
53	CLN	Citation Network	0.879	0.926	0.797	0.980	-0.107	0.064	-0.119	0.029
54	GD96	Citation Network	0.852	0.914	0.859	0.988	-0.283	0.099	-0.158	0.470
55	CGD	Citation Network	0.977	0.973	0.918	0.957	0.136	0.594	0.067	0.723
56	PDN	Citation Network	0.859	0.926	0.789	0.996	-0.203	-0.196	-0.115	-0.016
57	DLN	Literature Network	0.934	0.934	0.898	0.973	0.070	0.335	-0.068	0.320
58	GLN	Literature Network	0.918	0.914	0.832	0.922	-0.158	0.275	-0.197	0.664
59	PBN	Literature Network	0.980	0.969	0.918	0.992	-0.023	0.541	0.053	0.373
60	APN	Transportation Network	0.965	0.984	0.879	0.992	-0.207	-0.023	-0.149	0.055
61	LTN	Transportation Network	0.898	0.867	0.898	0.992	-0.209	0.626	-0.113	0.374
62	EUA	Transportation Network	0.941	0.980	0.801	0.996	-0.166	-0.005	-0.093	-0.084
63	ERD	Collaboration Network	0.980	0.977	0.957	0.973	0.182	0.403	0.047	0.569
64	MSJ	Collaboration Network	0.969	0.828	0.930	0.988	0.350	0.941	0.236	0.960
65	SWC	Game Network	0.867	0.918	0.816	0.973	0.080	0.140	0.026	0.108
66	FON	Game Network	0.992	0.988	0.965	0.996	0.191	0.693	0.059	0.313
67	MPN	Political Network	0.918	0.938	0.750	0.977	-0.155	0.132	-0.115	0.091
68	SPR	Political Network	0.930	0.949	0.922	0.988	0.019	0.140	-0.063	0.163
69	USS	Geographical Network	0.953	0.914	0.863	0.984	0.225	0.622	0.224	0.652
70	WTN	Trade Network	0.973	0.973	0.871	0.988	-0.016	0.019	-0.033	-0.029

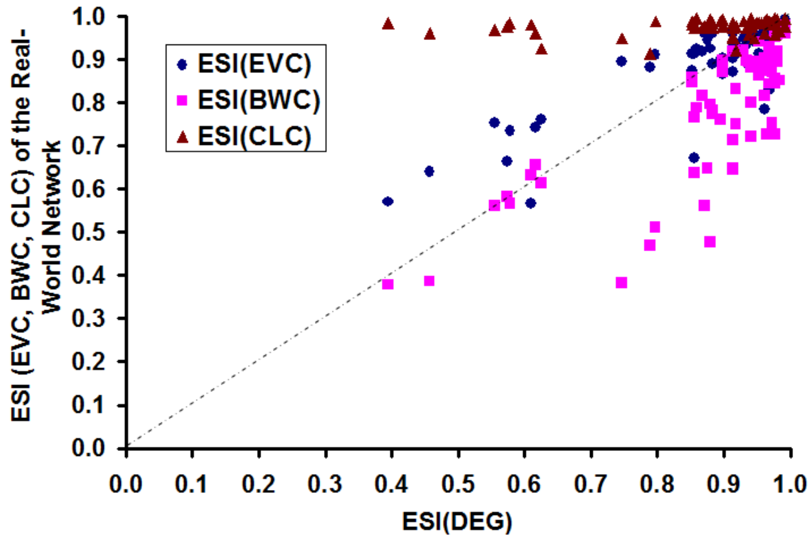
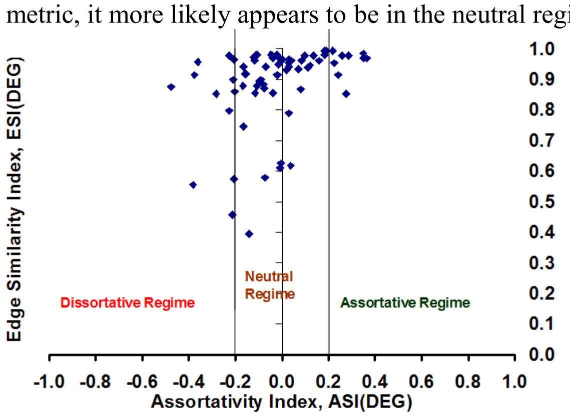


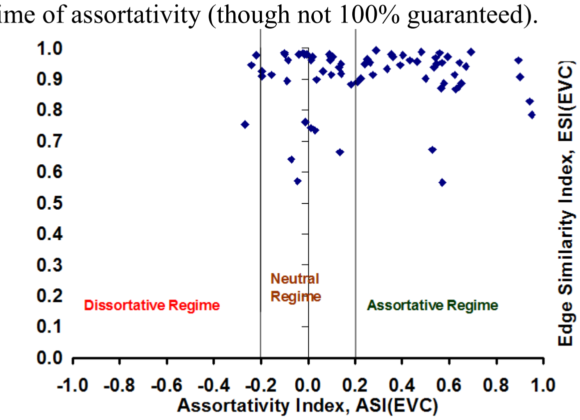
Figure 7: Comparison of the Centrality-based ESI Values for the Real-World Networks

Figure 7 plots the ESI(DEG) values vs. the ESI values with respect to the other three centrality metrics. The diagonal dotted line in the plot is used to identify the trend among the ESI values. If a data point is above the diagonal line, then the ESI with respect to the particular centrality metric is greater than ESI(DEG) and vice-versa. We observe the ESI(CLC) values to be far higher than that of the ESI(DEG) values, and there appears to be no correlation between the two. On the other hand, we observe at least a moderate level of correlation between the ESI(DEG) values vs. the ESI(EVC) and ESI(BWC) values. Considering all the 70 real-world networks, the Pearson's correlation coefficient between the ESI(DEG) and the ESI(EVC) values is 0.86 and the Pearson's correlation coefficient between the ESI(DEG) and the ESI(BWC) values is 0.77. Thus, using the ESI values of the real-world networks with respect to the computationally-light DEG centrality metric, we could predict the ESI values of the real-world networks with respect to the computationally-heavy EVC and BWC metrics.

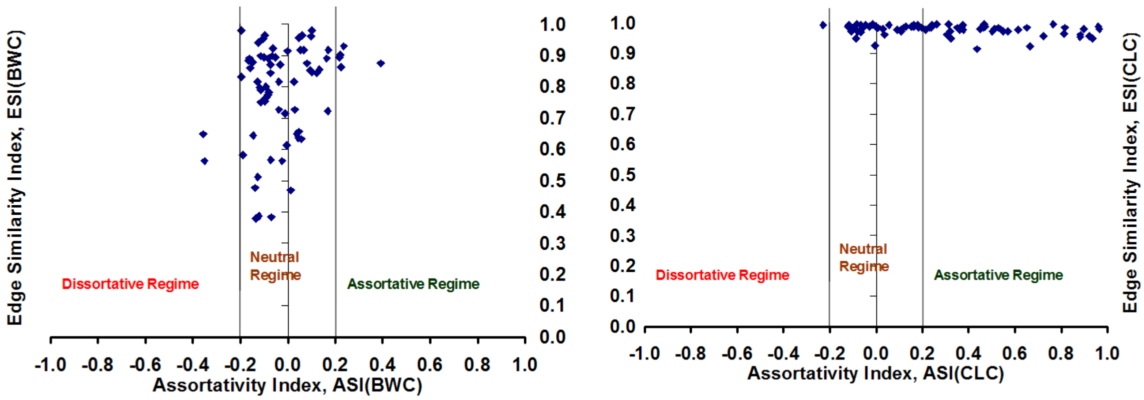
Figures 8.a-8.d illustrate the distribution of the ASI vs. ESI values for the four centrality metrics. In all the subfigures, we observe the real-world networks in any of the three regimes of assortativity to have comparable ESI values. In other words, it is not possible to assess the ESI value of a network-based on its ASI values. Several real-world networks in the neutral regime of assortativity incur a wide range of ESI values (especially, in the case of BWC) as well as incur predominantly larger ESI values (in the case of CLC, EVC and DEG). On the other hand, if a real-world network has a lower ESI value with respect to a centrality metric, it more likely appears to be in the neutral regime of assortativity (though not 100% guaranteed).



8.a: ASI (DEG) vs. ESI (DEG)



8.b: ASI (EVC) vs. ESI (EVC)



8.c: ASI (BWC) vs. ESI (BWC)   8.b: ASI (CLC) vs. ESI (CLC)

Figure 8: Comparison of the Centrality-based ESI Values for the Real-World Networks

Figure 9 plots the distribution of the spectral radius ratio for node degree (λsp) vs. the ESI values of the real-world networks with respect to the four centrality metrics. Except for CLC, we observe an inverse relationship between the λsp values and the ESI values for the other three centrality metrics. That is, a real-world network with a larger λsp value is more likely to have a lower ESI (DEG, EVC, BWC) compared to a real-world network with a lower λsp value. This implies, scale-free networks (that have larger λsp values) [16, 17] are more likely to incur lower ESI (DEG, EVC, BWC) values compared to the random networks (that have lower λsp values) [16, 18]. If a real-world network has a lower ESI with respect to DEG, EVC or BWC, it is more likely to have a larger λsp value (i.e., a larger variation in node degree). Such a trend is relatively stronger for BWC compared to DEG and EVC.

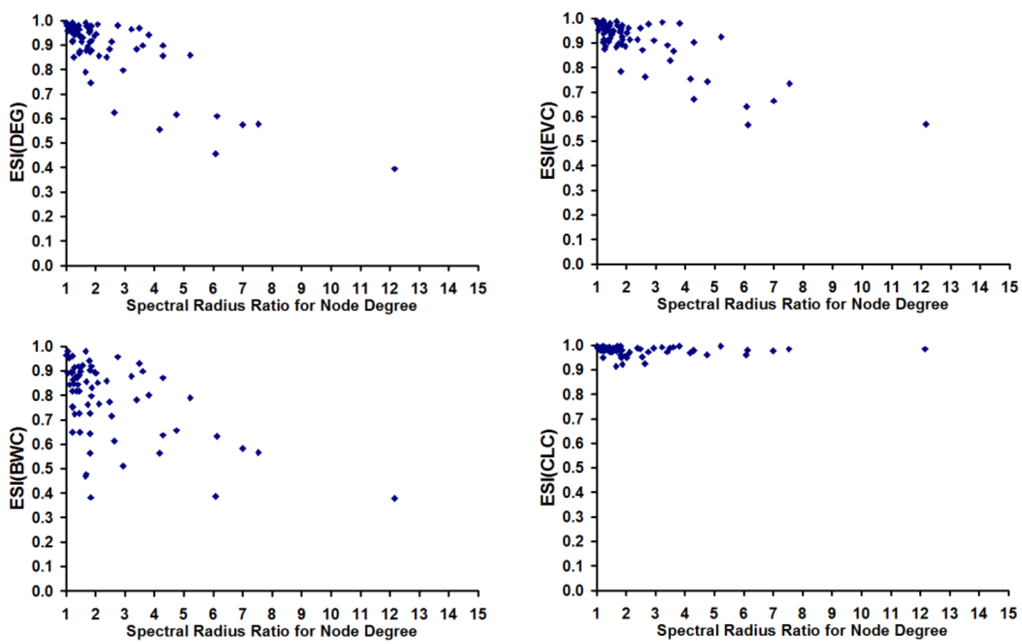


Figure 9: Spectral Radius Ratio for Node Degree vs. the ESI Values of the Real-World Networks

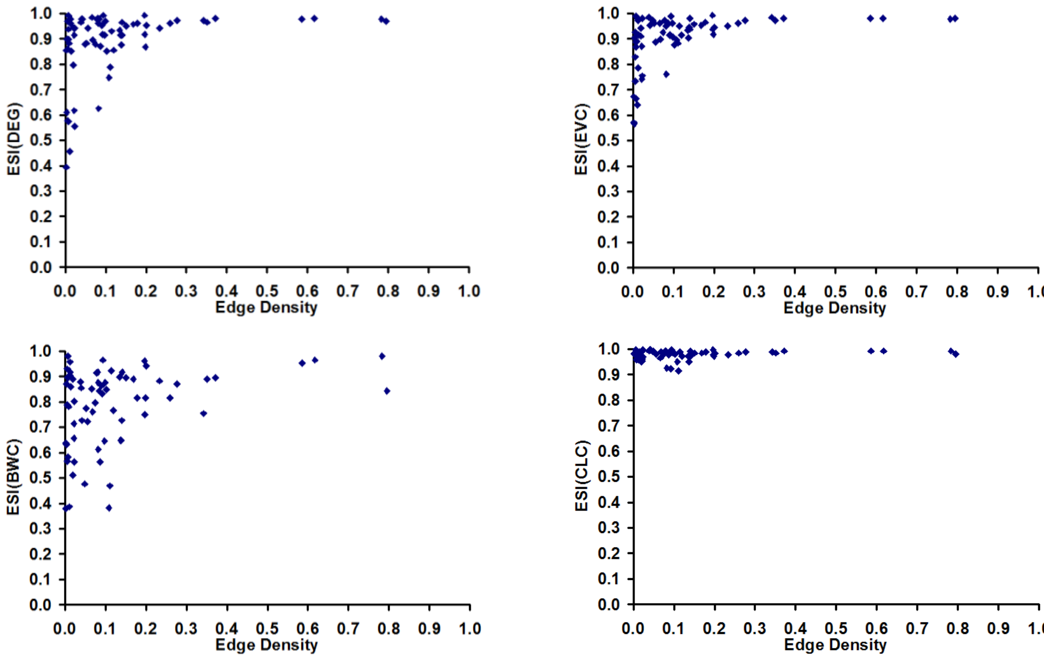


Figure 10: Edge Density vs. the ESI Values of the Real-World Networks

Figure 10 illustrates the distribution of the ESI values vs. the edge density (ρedge) of the real-world networks. With respect to any centrality metric, we observe real-world networks with a larger ρedge to incur a larger ESI value. Also, except for CLC, if a real-world network incurs a lower ESI value with respect to a centrality metric, it is more likely to have a lower ρedge. But, the converse is not necessarily true. We observe several real-world networks with lower ρedge values to incur significantly larger ESI values with respect to any of the four centrality metrics.

To further corroborate our observations in Figures 9-10 and assert the relationship between the ESI values and the spectral radius ratio for node degree as well as edge density, we compare (see Figure 11) the ESI(DEG) and ESI(BWC) values incurred for the 15 biological networks and the 12 acquaintance networks as well as the 11 friendship networks (grouped together as social networks). For both DEG and BWC, we observe the ESI values for the biological networks to be appreciably lower than the ESI values of the social networks. We observe the social networks to have a relatively larger edge density and lower spectral radius ratio for node degree: such real-world networks are more likely to incur larger ESI values (as seen in Figures 9 and 10). On the other hand, most of the biological networks are observed to have a relatively lower edge density and larger spectral radius ratio for node degree, both of which contributing to the lower ESI values for the biological networks.

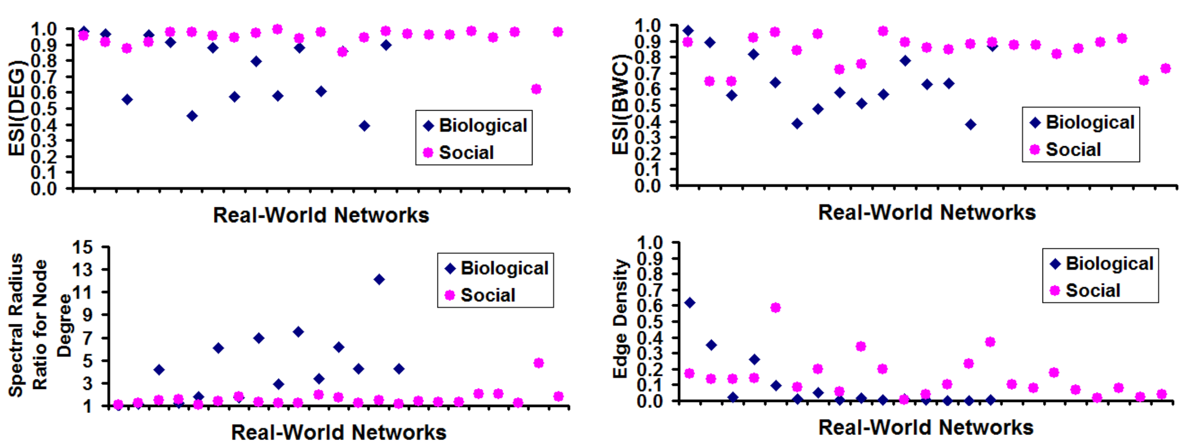


Figure 11: Biological Networks vs. Social Networks: ESI(DEG) and ESI(BWC) and their Relation to the Spectral Radius Ratio for Node Degree and Edge Density

5 Conclusions

The high-level contribution of this paper is the proposal of the Edge Similarity Index (ESI) measure to assess the similarity between any two edges on the basis of the node-level metric values of their end vertices. To the best of our knowledge, the ESI metric is the only quantitative network-level metric (in a scale of 0 to 1) to assess edge similarity in complex networks with respect to any node-level metric. Also, ours is the first approach to assess similarity on the basis of a logical coordinate system of the normalized node-level metric values for the end vertices of the edges. Another significant contribution of our research is the proposal of a binary search algorithm whose number of iterations is simply dependent on the cutoff parameter ($\in$) used as the basis to terminate the algorithm and is independent of the number of vertices and edges in the complex network. The overall time complexity of the algorithm is O(E2* $\log_2^{\sqrt{2}/\in}$), where E is the number of edges in the complex network and O(E2) is the time complexity to construct the unit-disk graph of the logical vertices (actual edges in the network) for an iteration.

The ESI metric is different from the assortativity index (ASI) metric in the sense that the latter quantifies the extent of similarity between the end vertices of any edge (and not any two edges, which is the basis for ESI) with respect to any node-level metric. We evaluate the ESI and ASI values for a suite of 70 real-world networks of diverse degree distributions and edge density with respect to four prototypical centrality metrics (neighbourhood-based degree: DEG and eigenvector: EVC; and shortest path-based betweenness: BWC and closeness: CLC) as the node-level metrics. We observe the ESI and ASI values for the real-world networks with respect to any of the four centrality metrics to be independent of each other. Both neutral as well as assortative/dissortative networks could incur larger ESI values. The only relationship we observe between the ESI and ASI metrics is that we observe real-world networks that incurred lower ESI values for a centrality metric (typically, BWC) to be typically neutral with respect to assortativity of the edges with respect to the centrality metric. But, it is very important to note that the converse need not be true.

We observed the ESI(CLC) values to be the largest for all the 70 real-world networks. We also observed the ESI(BWC) values to be relatively the lowest for a majority of the real-world networks. Though both BWC and CLC are shortest path-based centrality metrics, it is interesting to observe such a contrasting difference between their ESI values for any real-world network. This could be attributed to the fact that BWC could be viewed as a medial centrality metric [76] (i.e., captures the volume/number of shortest paths between any two vertices going through a particular vertex) whereas CLC could be viewed as a radial centrality metric [76] (i.e., captures the lengths of the shortest paths originating at a vertex). We also observed the ESI(DEG) values to exhibit a moderate level of linear correlation with the ESI(BWC) values and a strong level of linear correlation with the ESI(EVC) values. Hence, it is possible to use the ESI values of the real-world networks with respect to the computationally-light DEG metric as the basis to predict the ESI values with respect to the computationally-heavy BWC and EVC metrics.

We observe the ESI values of the real-world networks with respect to DEG, EVC and BWC to be related to the spectral radius ratio for node degree and edge density. Networks with larger spectral radius ratio for node degree (typically, such real-world networks are said to be scale-free in nature) incur lower ESI values, and networks with lower spectral radius ratio for node degree (typically, such real-world networks are said to be similar to random networks) incur larger ESI values. On the other hand, networks with larger edge density are observed to incur larger ESI values. We further corroborate our above claim by comparing the biological networks vs. social networks (includes both acquaintance and friendship networks). The biological networks are observed to have a relatively larger spectral radius ratio for node degree and the social networks are observed to have a relatively larger edge density. For both DEG and BWC, we observe the ESI values of the biological networks to be appreciably lower than the ESI values of the social networks. As part of future work, we plan to use the ESI measure as the basis to evaluate the similarity between any two real-world networks.

References

[1] Singhal, A. (2001). Modern information retrieval: A brief overview. *IEEE Data Eng. Bull.*, *24*(4), 35-43.
[2] Meghanathan, N. (2016). Maximal assortative matching and maximal dissortative matching for complex network graphs. *The Computer Journal*, *59*(5), 667-684.
[3] Newman, M.E.J. *Networks: An Introduction*. Oxford University Press, Oxford, UK, 2010.
[4] Newman, M.E.J. (2002). Assortative Mixing in Networks. Physical Review Letters, 89(20), 208701.
[5] Strang, G. (2005) Linear Algebra and its Applications. Brooks Cole, Pacific Grove, CA, USA
[6] Peterson, L. L., & Davie, B. S. (2007). *Computer networks: a systems approach*. Elsevier.
[7] McPherson, M., Smith-Lovin, L., & Cook, J.M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, *27*(1), 415-444.
[6] Bonacich, P. (1987). Power and centrality: A family of measures. *American journal of sociology*, *92*(5), 1170-1182.
[7] Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35-41.
[8] Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, *25*(2), 163-177.
[9] Freeman, L.C. (1978). Centrality in social networks conceptual clarification. *Social networks*, *1*(3), 215-239.
[10] Leiserson, C.E., Rivest, R.L., Cormen, T.H., & Stein, C. (2001). *Introduction to algorithms* (Vol. 6). Cambridge, MA: MIT press.
[11] Borgatti, S.P., & Everett, M.G. (2006). A graph-theoretic perspective on centrality. *Social networks*, *28*(4), 466-484.
[12] Meghanathan, N. (2016). Assortativity Analysis of Real-World Network Graphs based on Centrality Metrics. *Computer and Information Science*, *9*(3), 7-25.
[13] Piraveenan, M., Prokopenko, M., & Zomaya, A. Y. (2008). Local assortativeness in scale-free networks. *EPL (Europhysics Letters)*, *84*(2), 28002.
[14] Faudree, R.J., & Sheehan, J. (1998). The maximum number of edges in a graph with fixed edge-degree. *Discrete mathematics*, *183*(1-3), 81-101.
[15] Tan, P-N., Steinbach, M. and Kumar, V. *Introduction to Data Mining*, Pearson Publishers, 1st Edition, May 2005.
[16] Meghanathan, N. (2014, December). Spectral radius as a measure of variation in node degree for complex network graphs. In *2014 7th International Conference on u-and e-Service, Science and Technology* (pp. 30-33). IEEE.
[17] Barabasi, A.L. *Network Science*, 1st Edition, Cambridge University Press, August 2016.
[18] Erdos, P. and Renyi, A. (1959). On Random Graphs I, *Publicationes Mathematicae*, 6, 290-297.
[19] Takahata, Y. (1991). Diachronic changes in the dominance relations of adult female Japanese monkeys of the Arashiyama B group. *The monkeys of Arashiyama. State University of New York Press, Albany*, 123-139.
[20] De Reus, M.A., & van den Heuvel, M.P. (2013). Rich club organization and intermodule communication in the cat connectome. *Journal of Neuroscience*, *33*(32), 12929-12939.
[21] De Domenico, M., Porter, M.A., & Arenas, A. (2015). MuxViz: a tool for multilayer analysis and visualization of networks. *Journal of Complex Networks*, *3*(2), 159-176.
[22] Ulanowicz, R.E., & DeAngelis, D.L. (2005). Network analysis of trophic dynamics in south florida ecosystems. *US Geological Survey Program on the South Florida Ecosystem*, *114*, 45.
[23] Zander, C. D., Josten, N., Detloff, K.C., Poulin, R., McLaughlin, J.P., & Thieltges, D.W. (2011). Food web including metazoan parasites for a brackish shallow water ecosystem in Germany and Denmark: Ecological Archives E092-174. *Ecology*, *92*(10), 2007-2007.

[24] White, J.G., Southgate, E., Thomson, J.N., & Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci*, *314*(1165), 1-340.
[25] De Domenico, M., Nicosia, V., Arenas, A., & Latora, V. (2015). Structural reducibility of multilayer networks. *Nature communications*, *6*, 6864.
[26] Duch, J. and Arenas, A. (2005). Communication Detection in Complex Networks using Extremal Optimization, *Physical Review E* 72, 027104.
[27] Bhardwaj, N., Yan, K.K., & Gerstein, M.B. (2010). Analysis of diverse regulatory networks in a hierarchical context shows consistent tendencies for collaboration in the middle levels. *Proceedings of the National Academy of Sciences*, *107*(15), 6841-6846.
[28] Arabidopsis Interactome Mapping Consortium. (2011). Evidence for network evolution in an Arabidopsis interactome map. *Science*, *333*(6042), 601-607.
[29] De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M.A., Gómez, S. & Arenas, A. (2013). Mathematical formulation of multilayer networks. *Physical Review X*, *3*(4), 041022.
[30] Walker, D., Xie, H., Yan, K.K., & Maslov, S. (2007). Ranking scientific publications using a model of network traffic. *Journal of Statistical Mechanics: Theory and Experiment*, *2007*(06), P06010.
[31] Schwimmer, E. (1973). *Exchange in the social structure of the Orokaiva: traditional and emergent ideologies in the Northern District of Papua*. London: C. Hurst.
[32] Michael, J.H. (1997). Labor dispute reconciliation in a forest products manufacturing facility. *Forest products journal*, *47*(11/12), 41.
[33] Zachary, W.W. (1977). An information flow model for conflict and fission in small groups. *Journal of anthropological research*, *33*(4), 452-473.
[34] Rogers, E.M., and Kincaid, D.L. (1980). Communication Networks: Toward a New Paradigm for Research, Free Press, New York City, NY, USA.
[35] Bernard, H.R., Killworth, P.D., and 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.
[36] Lusseau, D., Schneider, K., Boisseau, O.J., Haase, P., Slooten, E., and Dawson, S.M. (2003). The Bottlenose Dolphin Community of Doubtful Sound Features a Large Proportion of Long-lasting Associations. *Behavioral Ecology and Sociobiology*, 54(3), 396-405.
[37] Hayes, B. (2006). Connecting the Dots. *American Scientist*, 94(5), 400-404.
[38] Loomis, C.P., Morales, J.O., Clifford, R.A., and Leonard, O. E. (1953). Turrialba Social Systems and the Introduction of Change, 45-78, The Free Press, Glence, IL, USA.
[39] Isella, L., Stehle, J., Barrat, A., Cattuto, C., Pinton, J.F., and Van den Broeck, W. (2011). What's in a Crowd? Analysis of Face-to-Face Behavioral Networks. *Journal of Theoretical Biology*, 271(1), 166-180.
[40] Gemmetto, V. Barrat, A., and Cattuto, C. (2014). Mitigation of Infectious Disease at School: Targeted Class Closure vs. School Closure. *BMC Infectious Diseases*, 14(695), 1-10.
[41] 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.
[42] Batagelj, V., and Mrvar, A. (2005). Pajek Datasets, http://vlado.fmf.uni-lj.si/pub/networks/data/.
[43] Krackhardt, D. (1999). The Ties that Torture: Simmelian Tie Analysis in Organizations. *Research in the Sociology of Organizations*, 16, 183-210.
[44] Freeman, L.C., Freeman, S.C., and Michaelson, A.G. (1989). How Humans See Social Groups: A Test of the Sailer-Gaulin Models. *Journal of Quantitative Anthropology*, 1, 229-238.
[45] Pearson, M., and Michell, L. (2000). Smoke Rings: Social Network Analysis of Friendship Groups, Smoking and Drug-taking. Drugs: Education. *Prevention and Policy*, 7(1), 21-37.

[46] MacRae, D. (1960). Direct Factor Analysis of Sociometric Data. *Sociometry*, 23(4), 360-371.
[47] Nepusz, T., Petroczi, A., Negyessy, L., and Bazso, F. (2008). Fuzzy Communities and the Concept of Bridgeness in Complex Networks. *Physical Review E*, 77(1), 016107.
[48] 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., and 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.
[49] Freeman, L.C., Webster, C.M., and Kirke, D.M. (1998). Exploring Social Structure using Dynamic Three-Dimensional Color Images. *Social Networks*, 20(2), 109-118.
[50] Coleman, J., Katz, E., & Menzel, H. (1957). The diffusion of an innovation among physicians. *Sociometry*, *20*(4), 253-270.
[51] Blagus, N., Šubelj, L., & Bajec, M. (2012). Self-similar scaling of density in complex real-world networks. *Physica A: Statistical Mechanics and its Applications*, *391*(8), 2794-2802.
[52] Knuth, D. E. (1993). *The Stanford GraphBase: a platform for combinatorial computing* (pp. 74-87). New York: AcM Press.
[53] Newman, M. (2006). Finding Community Structure in Networks using the Eigenvectors of Matrices. *Physics Review*, E 74, 036104.
[54] Batagelj, V. and Mrvar, A. *Pajek Datasets*, http://vlado.fmf.uni-lj.si/pub/networks/data/, 2006.
[55] Gleiser, P.M. (2007). How to become a Superhero. *Journal of Statistical Mechanics: Theory and Experiments*, P09020.
[56] Moreno, J.L. (1960). The Sociometry Reader. 534-547, The Free Press, Glence, IL, USA.
[57] Magnani, M., Micenkova, B., & Rossi, L. (2013). Combinatorial analysis of multiple networks. *arXiv preprint arXiv:1303.4986*.
[58] Lazega, E. (2001). *The collegial phenomenon: The social mechanisms of cooperation among peers in a corporate law partnership*. Oxford University Press on Demand.
[59] Cross, R.L., Parker, A., and Cross, R. (2004). The Hidden Power of Social Networks: Understanding How Work Really Gets Done in Organizations. Harvard Business Review Press, Watertown, MA, USA.
[60] Geiser, P., and Danon, L. (2003). Community Structure in Jazz. *Advances in Complex Systems*, 6(4), 563-573.
[61] Scott, J.P. (1980). The Anatomy of Scottish Capital: Scottish Companies and Scottish Capital, Croom Helm, London, UK.
[62] Hummon, N.P., Doreian, P., and Freeman, L.C. (1990). Analyzing the Structure of the Centrality-Productivity Literature Created between 1948 and 1979. *Science Communication*, 11(4), 459-480.
[63]Batagelj, V., and Mrvar, A. (2005). Pajek Datasets, http://vlado.fmf.uni-lj.si/pub/networks/data/.
[64] Biedl, T., and Franz, B.J. (2001). Graph-Drawing Contest Report. Proceedings of the 9th International Symposium on Graph Drawing, Austria, Vienna, 23-26 September, pp.~513-521, Springer, USA.
[65] Heymann, S. CPAN-Explorer, an Interactive Exploration of the Perl Ecosystem, Gephi Blog, 2009.
[66] Nooy, W. (1999). A Literary Playground: Literary Criticism and Balance Theory. *Poetics*, 26(5-6), 385-404.
[67] Krebs, V. (2003). Proxy Networks: Analyzing One Network to Reveal Another. *Bulletin de Méthodologie Sociologique*, 79, 61-40.
[68] De Domenico, M., Sole-Ribalta, A., Gomez, S. and Arenas, A. Navigability of Interconnected Networks under Random Failures, *Proceedings of the National Academy of Sciences*, vol. 111, pp. 8351-8356, 2014.

[69] Cardillo, A., Gómez-Gardenes, J., Zanin, M., Romance, M., Papo, D., Del Pozo, F., & Boccaletti, S. (2013). Emergence of network features from multiplexity. *Scientific reports*, *3*, 1344.
[70] McCarty, C., and Freeman, L. (2008). http://moreno.ss.uci.edu/data.html.
[71] Girvan, M., and Newman, M.E.J. (2002). Community Structure in Social and Biological Networks. Proceedings of the National Academy of Sciences of the United States of America, 99(12), 7821-7826.
[72] Gil-Mendieta, J., and Schmidt, S. (1996). The Political Network in Mexico. *Social Networks*, 18(4), 355-381.
[73] Grimmer, J. (2010). A Bayesian Hierarchical Topic Mode for Political Texts: Measuring Expressed Agendas in Senate Press Releases. *Political Analysis*, 18(1), 1-35.
[74] Meghanathan, N. (2017). Complex Network Analysis of the Contiguous United States Graph. *Computer and Information Science*, *10*(1), 54-76.
[75] Smith, D.A., and White, D.R. (1992). Structure and Dynamics of the Global Economy: Network Analysis of International Trade 1965-1980. *Social Forces*, 70(4), 857-893.
[76] Borgatti, S.P., & Everett, M.G. (2006). A graph-theoretic perspective on centrality. *Social networks*, *28*(4), 466-484.