Computer and Information Science; Vol. 10, No. 1; 2017
ISSN 1913-8989 E-ISSN 1913-8997
Published by Canadian Center of Science and Education

# Complex Network Analysis of the Contiguous United States Graph

Natarajan Meghanathan[1]

[1] 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: December 20, 2016 Accepted: January 29, 2017 Online Published: February 5, 2017

doi:10.5539/cis.v10n1p54 URL: http://dx.doi.org/10.5539/cis.v10n1p54

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

**Abstract**

We model the contiguous states (48 states and the District of Columbia) of the United States (US) as an undirected network graph with each state represented as a node and there is an edge between two nodes if the corresponding two states share a common border. We determine a ranking of the states in the US with respect to a suite of node-level metrics: the centrality metrics (degree, eigenvector, betweenness and closeness), eccentricity, maximal clique size, and local clustering coefficient. We propose a normalization-based approach to obtain a comprehensive centrality ranking of the vertices (that is most likely to be tie-free) encompassing the normalized values of the four centrality metrics. We have applied the proposed normalization-based approach on the US States graph to obtain a tie-free ranking of the vertices based on a comprehensive centrality score. We observe the state of Missouri to be the most central state with respect to all the four centrality metrics. We have also analyzed the US States graph with respect to a suite of network-level metrics: bipartivity index, assortativity index, modularity, size of the minimum connected dominating set, algebraic connectivity and degree metrics. The approach taken in this paper could be useful for several application domains: transportation networks (to identify central hubs), politics (to identify campaign venues with larger geographic coverage), cultural and electoral studies (to identify communities of states that are relatively proximal to each other) and etc.

**Keywords:** Network Analysis, Centrality, Clique, Bipartivity, Modularity

## 1. Introduction

Network Science is one of the emerging fields of Data Science to analyze real-world networks from a graph theory point of view. Several real-world networks have been successfully modeled as undirected and directed graphs to study the intrinsic structural properties of the networks as well as the topological importance of nodes in these networks. The real-world networks that have been subjected to complex network analysis typically fall under one of these categories: social networks (Ghali et al., 2012), transportation networks (Cheung & Gunes, 2012), biological networks (Ma & Gao, 2012), citation networks (Zhao & Strotmann, 2015), co-authorship networks (Ding, 2011) and etc. One category of real-world networks for which sufficient attention has not yet been given are the regional networks featuring the states within a country.

In this paper, we present a comprehensive analysis of a network graph of the states within a country with respect to various node-level and network-level metrics typically considered in the field of Network Science and demonstrate the utility of information that can be obtained from the analysis. We also propose a normalization-based approach to obtain comprehensive centrality scores for the vertices encompassing the normalized individual centrality scores and illustrate the use of these comprehensive scores to obtain a ranking of the vertices (that is most likely to be tie-free). We also illustrate the procedure to identify the centrality metric whose scores and ranking are relatively the closest to the normalized comprehensive centrality scores and ranking.

We opine the paper to serve as a model for anyone interested in analyzing a connected graph of the states within a country from a Network Science perspective. The approaches presented in this paper could be useful to determine the states (and their cities) that are the most central and/or influential within a country. For example, the ranking of the vertices based on the shortest path centrality metrics (closeness and betweenness) could be useful to choose the states (and their cities) that could serve as hubs for transportation networks (like road and airline networks). We could identify the states that are most the central states as well as identify the states that could form a connected backbone and geographically well-connected to the rest of the states within a country and use this information to design the road/rail transportation networks. The degree centrality and eigenvector centrality metrics as well as the network-level metrics like minimum connected dominating set and maximal clique size could be useful to identify fewer number of venues (with several adjacent states to draw people) for political campaigns/meetings that would cover the entire country. Node-level metrics like local clustering coefficient could be useful to identify the states that are critical to facilitate communication between the neighbor states. One could develop an optimal regional classification of states for cultural studies (language accent, eating habits, etc) and electoral studies (like scheduling of elections) by identifying communities of states (that are relatively more proximal with each other) with high modularity scores.

Table 1. List of Contiguous States (including DC) of the US in Alphabetical Order

| ID | State/District | Code | ID | State/District | Code |
|---|---|---|---|---|---|
| 1 | Alabama | AL | 26 | Nebraska | NE |
| 2 | Arizona | AZ | 27 | Nevada | NV |
| 3 | Arkansas | AR | 28 | New Hampshire | NH |
| 4 | California | CA | 29 | New Jersey | NJ |
| 5 | Colorado | CO | 30 | New Mexico | NM |
| 6 | Connecticut | CT | 31 | New York | NY |
| 7 | Delaware | DE | 32 | North Carolina | NC |
| 8 | District of Columbia | DC | 33 | North Dakota | ND |
| 9 | Florida | FL | 34 | Ohio | OH |
| 10 | Georgia | GA | 35 | Oklahoma | OK |
| 11 | Idaho | ID | 36 | Oregon | OR |
| 12 | Illinois | IL | 37 | Pennsylvania | PA |
| 13 | Indiana | IN | 38 | Rhode Island | RI |
| 14 | Iowa | IA | 39 | South Carolina | SC |
| 15 | Kansas | KS | 40 | South Dakota | SD |
| 16 | Kentucky | KY | 41 | Tennessee | TN |
| 17 | Louisiana | LA | 42 | Texas | TX |
| 18 | Maine | ME | 43 | Utah | UT |
| 19 | Maryland | MD | 44 | Vermont | VT |
| 20 | Massachusetts | MA | 45 | Virginia | VA |
| 21 | Michigan | MI | 46 | Washington | WA |
| 22 | Minnesota | MN | 47 | West Virginia | WV |
| 23 | Mississippi | MS | 48 | Wisconsin | WI |
| 24 | Missouri | MO | 49 | Wyoming | WY |
| 25 | Montana | MT | | | |

We choose the United States (US) as the country for analysis and build a connected network graph of the contiguous states (48 states and the District of Columbia, DC) of the US: each state and DC is a node (vertex) and there exists a link (edge) between two vertices if the two corresponding states/DC share a common border. Though some prior studies have been conducted on transportation networks (Cheung & Gunes, 2012) and food flow networks (Lin et al., 2014) in the United States, to the best of our knowledge, there has been no prior study of network analysis on the graph of the contiguous US states solely based on their geographical locations. In this paper, we have implemented the algorithms to compute several node-level metrics (such as the degree centrality, eigenvector centrality (Newman, 2010), betweenness centrality (Brandes, 2001), closeness centrality (Newman, 2010), maximal clique size (Meghanathan, 2015b), eccentricity (Cormen et al., 2009) and local clustering coefficient) as well as several network-level metrics (such as bipartivity index (Estrada & Rodriguez-Velazquez, 2005), modularity (Newman, 2006), minimum connected dominating set (Meghanathan, 2014b), algebraic connectivity (Fiedler, 1973), average path length (Cormen et al., 2009), diameter (Cormen et al., 2009), assortativity index (Newman, 2010) and spectral radius (Meghanathan, 2014a)) and analyze the US States network graph with respect to these metrics. We also analyze random network instances (generated with the same degree sequence using the Configuration model (Meghanathan, 2016c)) of the US States graph to study the correlation of the node-level metrics and proximity of values for the network-level metrics. Finally, we illustrate the application of the proposed normalized comprehensive centrality (NCC) scores-based ranking of the vertices on the US States network graph and observe that the NCC-based ranking of the vertices is indeed tie-free for the graph. We also identify the eigenvector centrality metric to be the centrality metric whose normalized scores and ranking of the vertices have relatively the lowest RMSD (root mean square difference) value to that of the NCC scores and the NCC-based ranking of the vertices.

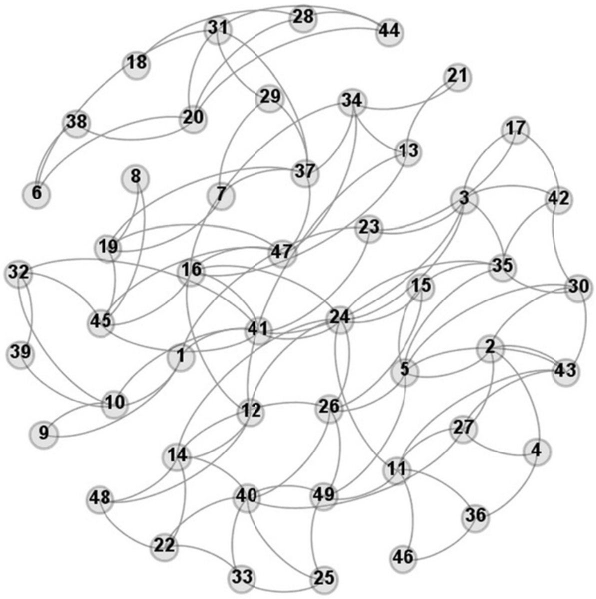


Figure 1. Fruchterman Reingold Layout of the US Network States Graph

Table 1 lists the contiguous states and DC in alphabetical order, their two character codes and the IDs used to refer to them in the paper. The rest of the paper is organized as follows: Section 2 introduces the node-level metrics and presents the results of the analysis on the states graph for each of them. Section 3 introduces the network-level metrics and presents the results of analysis on the states graph for each of them. Section 4 presents the normalization-based approach to obtain a comprehensive centrality ranking of the vertices and its application on the US States graph. Section 5 analyzes the random network instances (generated with the same degree sequence using the Configuration model) of the US States graph and compares the correlation of the node-level metrics and proximity of the network-level metrics. Section 6 discusses related work. Section 7 concludes the paper by summarizing the results of Sections 2-5. For the rest of the paper, the terms 'network' and 'graph', 'node' and 'vertex', 'link' and 'edge' are used interchangeably. They mean the same. The layout for the US States Network graph presented in Figure 1 is drawn using the Fructherman Reingold layout algorithm (Fruchterman & Reingold, 1991), available in Gephi (Cherven, 2015).

## 2. Node-Level Metrics

In this section, we introduce the node-level metrics for which we will run their respective algorithms on the US States Network graph and present the results (including their distribution and ranking of the vertices). The node-level metrics discussed include the four centrality metrics (degree centrality, eigenvector centrality, closeness centrality and betweenness centrality), maximal clique size, local clustering coefficient and the distance metrics: path length, eccentricity and radius. Since the US States graph is an undirected graph, the adjacency matrix of the graph is symmetric and there is only one value per vertex for each node-level metric.

### *2.1 Degree Centrality*

The degree centrality (DegC) of a vertex is the number of edges incident on it. Table 2 presents the degree centrality of the vertices and the corresponding rank (in the decreasing order of their values) in the US States Network graph; vertices with identical values for DegC have the same rank. The state of Missouri has the largest degree centrality value of 9, followed by the state of Tennessee with the second largest degree centrality value of 8. The state of Maine has the smallest degree centrality value of 1 (as New Hampshire is its only adjacent state). There are no ties among vertices for the largest and second largest values of degree centrality as well as for the vertex with the smallest degree centrality. However, as we can notice (from Table 2): for other values of degree centrality, there are several instances of ties among vertices (we assign the same rank for all such vertices with identical values for degree centrality).

Table 2. Ranking of the Vertices in the US States Network Graph based on Degree Centrality (DegC)

| Rank | ID | DegC | Rank | ID | DegC | Rank | ID | DegC | Rank | ID | DegC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 24 | 9 | 5 | 2 | 5 | 6 | 23 | 4 | 7 | 48 | 3 |
| 2 | 41 | 8 | 5 | 30 | 5 | 6 | 42 | 4 | 7 | 28 | 3 |
| 3 | 5 | 7 | 5 | 27 | 5 | 6 | 15 | 4 | 7 | 44 | 3 |
| 3 | 16 | 7 | 5 | 43 | 5 | 6 | 32 | 4 | 7 | 33 | 3 |
| 3 | 40 | 7 | 5 | 49 | 5 | 6 | 13 | 4 | 7 | 25 | 3 |
| 4 | 3 | 6 | 5 | 20 | 5 | 6 | 22 | 4 | 8 | 9 | 2 |
| 4 | 35 | 6 | 5 | 31 | 5 | 7 | 4 | 3 | 8 | 38 | 2 |
| 4 | 26 | 6 | 5 | 19 | 5 | 7 | 17 | 3 | 8 | 8 | 2 |
| 4 | 37 | 6 | 5 | 12 | 5 | 7 | 36 | 3 | 8 | 39 | 2 |
| 4 | 45 | 6 | 5 | 34 | 5 | 7 | 6 | 3 | 8 | 46 | 2 |
| 4 | 11 | 6 | 5 | 47 | 5 | 7 | 7 | 3 | 8 | 21 | 2 |
| 4 | 14 | 6 | 6 | 1 | 4 | 7 | 29 | 3 | 9 | 18 | 1 |
| 5 | 10 | 5 | | | | | | | | | |

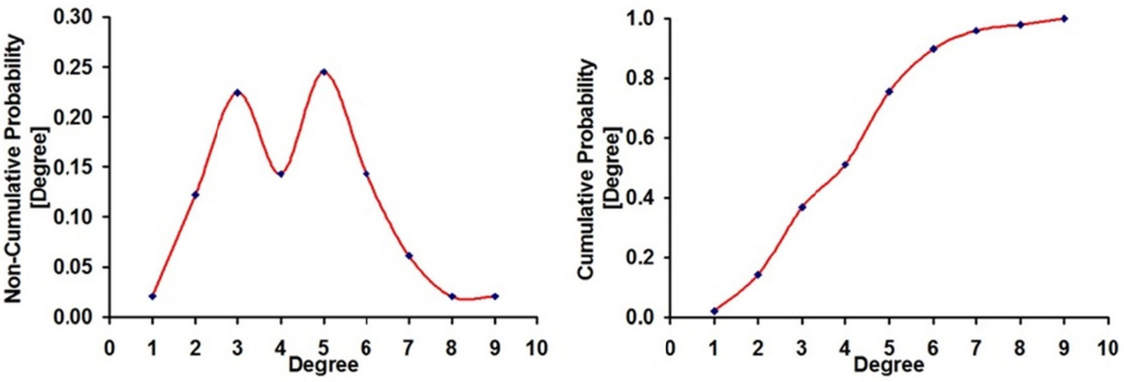


Figure 2. Bi-Modal Poisson Distribution (Non-Cumulative and Cumulative) for the Degree Centrality Metric of the US States Graph

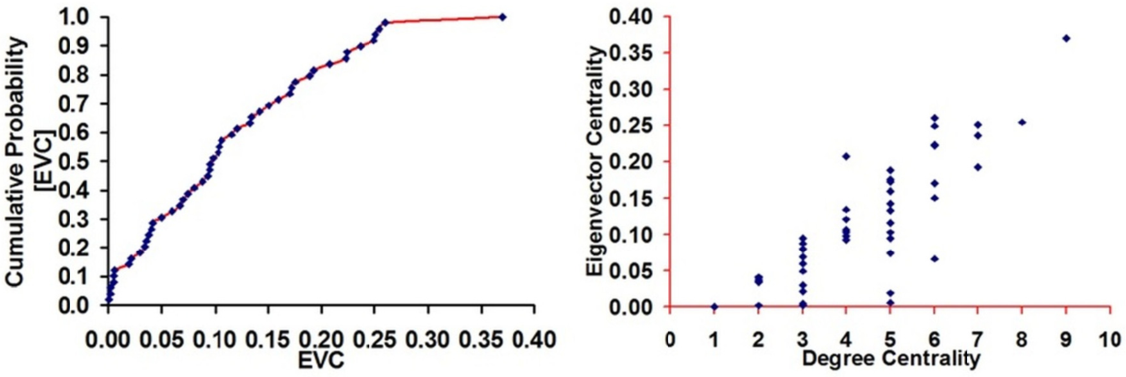
Figure 2 illustrates that the non-cumulative and cumulative degree distributions of the vertices in the states graph. The non-cumulative distribution curve illustrates that the degree distribution is bi-modal with a Poisson pattern (Balakrishnan & Nevzorov, 2003) with the peaks observed at degree values of 3 and 5; the cumulative distribution curve indicates that more than 85% of the vertices have degree of 6 or lower (the degree values are mostly ± 1 away from either of the two peaks), even though the largest degree observed is 9.

### *2.2 Eigenvector Centrality*

The eigenvector centrality (EVC) of a vertex is a measure of the degree of the vertex as well as the degree of its neighbors. A vertex is likely to have a higher EVC if it has a larger degree and its neighbor(s) also have a larger degree. The EVC for a vertex corresponds to the entry for the vertex in the principal eigenvector of the adjacency matrix of the graph (Chung, 2006). The EVC of the vertices is computed by implementing the power-iteration algorithm (Lay et al., 2015). Table 3 lists the EVC values (rounded to four decimal places) of the vertices in the US States graph. We observe a tie between the states of Vermont and Connecticut (values of 0.0050 each). The rest of the 47 vertices have a unique ranking. The US States graph is another example to illustrate that (unlike the degree centrality metric) the eigenvector centrality metric is more likely to return unique values for the vertices in real-world network graphs (Meghanathan, 2015a). The state of Missouri has the largest EVC value (0.3697), distantly followed by the states of Nebraska (0.2605) and Tennessee (0.2546); the state of Maine has the lowest EVC value (0.0004), far away from the immediately larger value of 0.0020 for the state of Rhode Island. The largest values for the state of Missouri with respect to both the degree centrality and eigenvector centrality metrics indicate that the state of Missouri not only has the largest degree, its neighboring states also have a relatively larger degree.

Table 3. Ranking of the Vertices in the US States Network Graph based on Eigenvector Centrality (EVC)

| Rank | ID | EVC | Rank | ID | EVC | Rank | ID | EVC | Rank | ID | EVC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 24 | 0.3697 | 14 | 11 | 0.1703 | 26 | 48 | 0.0951 | 38 | 21 | 0.0380 |
| 2 | 26 | 0.2605 | 15 | 43 | 0.1595 | 27 | 10 | 0.0946 | 39 | 39 | 0.0354 |
| 3 | 41 | 0.2546 | 16 | 45 | 0.1506 | 28 | 1 | 0.0927 | 40 | 9 | 0.0344 |
| 4 | 5 | 0.2509 | 17 | 2 | 0.1423 | 29 | 17 | 0.0880 | 41 | 7 | 0.0298 |
| 5 | 35 | 0.2494 | 18 | 42 | 0.1346 | 30 | 25 | 0.0803 | 42 | 29 | 0.0212 |
| 6 | 16 | 0.2362 | 19 | 27 | 0.1330 | 31 | 19 | 0.0742 | 43 | 31 | 0.0191 |
| 7 | 3 | 0.2235 | 20 | 23 | 0.1210 | 32 | 33 | 0.0696 | 44 | 20 | 0.0061 |
| 8 | 14 | 0.2227 | 21 | 47 | 0.1158 | 33 | 37 | 0.0666 | 45 | 44 | 0.0050 |
| 9 | 15 | 0.2076 | 22 | 22 | 0.1065 | 34 | 4 | 0.0597 | 45 | 6 | 0.0050 |
| 10 | 40 | 0.1924 | 23 | 13 | 0.1039 | 35 | 36 | 0.0496 | 46 | 28 | 0.0021 |
| 11 | 12 | 0.1887 | 24 | 34 | 0.1029 | 36 | 8 | 0.0413 | 47 | 38 | 0.0020 |
| 12 | 49 | 0.1752 | 25 | 32 | 0.0983 | 37 | 46 | 0.0404 | 48 | 18 | 0.0004 |
| 13 | 30 | 0.1720 | | | | | | | | | |



(a) Cumulative Probability Distribution of EVC (b) Degree Centrality vs. Eigenvector Centrality

Figure 3. Eigenvector Centrality of the Vertices in the US States Graph

Figure 3-a illustrates the cumulative probability distribution of the EVC values of the vertices in the US States graph. Except the state of Missouri (that has an EVC value much larger than the rest of the vertices), each of the other states have an EVC value that is closer to one or two other states. Thus, the distribution of the EVC values of the vertices follows a Poisson distribution. Figure 3-b presents a comparison of the degree centrality and eigenvector centrality values of the vertices and the Spearman's rank-based correlation coefficient (Daniel, 2000) between the two degree-based centrality metrics is 0.80. We observe that vertices with the same degree centrality have a wide range of values for the eigenvector centrality. Though vertices with larger degree centrality appear to be more likely to have a larger eigenvector centrality, there are several vertices for which the eigenvector centrality is relatively lower (compared to the EVC of vertices that have a relatively lower degree centrality) even if they have a higher degree centrality. For example, the state of Massachusetts (degree centrality - 5) has a much lower EVC (0.0061) than the state of Georgia (degree centrality - 3; EVC - 0.0946): primarily, attributed to the relatively higher degree of the neighbors for the state of Georgia.

### *2.3 Betweenness Centrality*

The Betweenness Centrality (BWC) of a vertex is a measure of its presence among the shortest paths between any two vertices in the graph. The BWC of a vertex $v_i$ is the sum of the fractions of the shortest paths between any two vertices ($v_j$ and $v_k$; $i \neq j \neq k$) that go through $v_i$. Quantitatively,

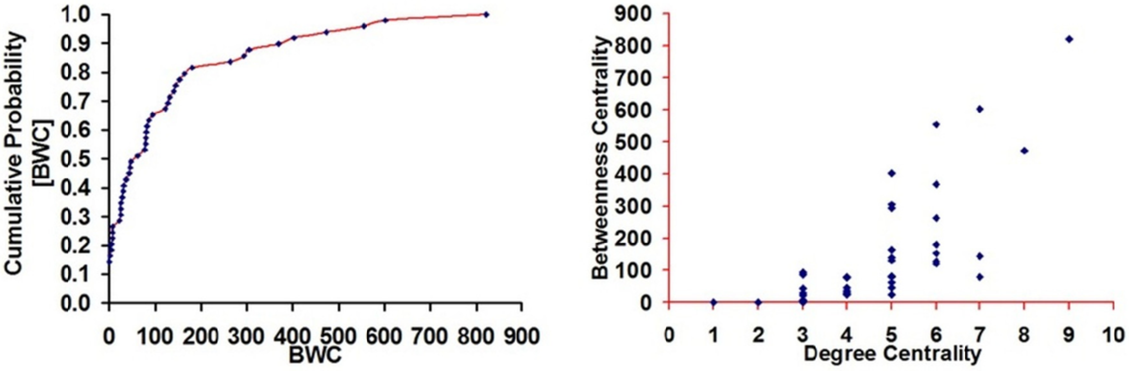
$$BWC(i) = \sum_{j \neq k \neq i} \frac{sp_{jk}(i)}{sp_{jk}},$$

where $sp_{jk}$ is the total number of shortest paths between vertices $v_j$ and $v_k$; $sp_{jk}$ $(i)$ is the number of such shortest paths between $v_j$ and $v_k$ that go through vertex $v_i$. We determine the BWC of the vertices in the US States graph by implementing the Brandes' algorithm (Brandes, 2001). Table 4 ranks the vertices in the US States graph based on BWC; the state of Missouri has the largest BWC value (821.4), followed by the states of Kentucky (602.4) and Pennsylvania (554.4). Seven vertices (DC, Florida, Maine, Michigan, Rhode Island, South Carolina and Washington) have a BWC value of 0.0 - indicating that these vertices do not lie on the shortest path between any two vertices in the graph. In addition to the above tie, when rounded to the first decimal value for the non-zero values of BWC, we notice that there are ties between Delaware and New Jersey (with a lower BWC of 7.0 each). A total of 42 distinct rank values could be assigned for the vertices.

Table 4. Ranking of the Vertices in the US States Network Graph based on Betweenness Centrality (BWC)

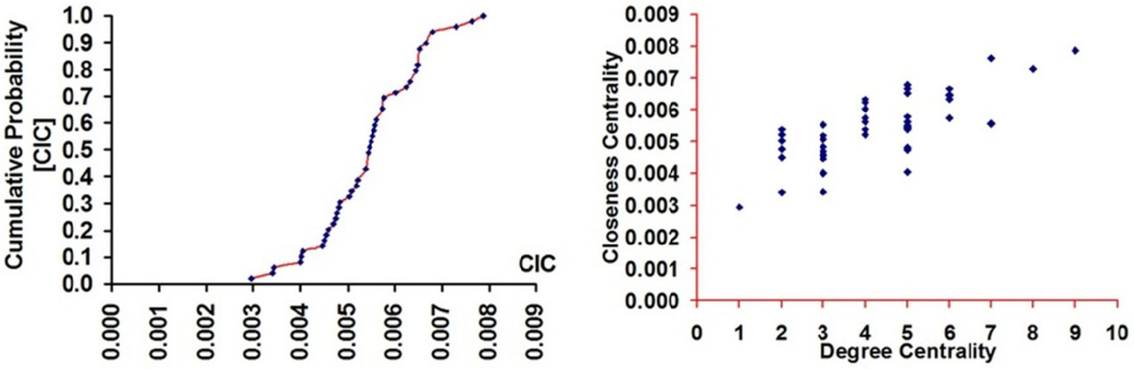
| Rank | ID | BWC | Rank | ID | BWC | Rank | ID | BWC | Rank | ID | BWC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 24 | 821.4 | 14 | 20 | 140.0 | 26 | 1 | 47.0 | 38 | 7 | 7.0 |
| 2 | 16 | 602.4 | 15 | 19 | 131.0 | 27 | 30 | 45.5 | 38 | 29 | 7.0 |
| 3 | 37 | 554.4 | 16 | 26 | 127.2 | 28 | 6 | 43.0 | 39 | 17 | 4.7 |
| 4 | 41 | 472.7 | 17 | 35 | 121.4 | 29 | 22 | 36.7 | 40 | 25 | 4.5 |
| 5 | 31 | 403.0 | 18 | 28 | 94.0 | 30 | 23 | 31.3 | 41 | 33 | 1.7 |
| 6 | 11 | 368.9 | 19 | 44 | 86.0 | 31 | 36 | 30.4 | 42 | 8 | 0.0 |
| 7 | 34 | 305.6 | 20 | 27 | 81.8 | 32 | 42 | 28.1 | 42 | 9 | 0.0 |
| 8 | 47 | 294.1 | 21 | 5 | 80.4 | 33 | 2 | 25.4 | 42 | 18 | 0.0 |
| 9 | 45 | 264.4 | 22 | 10 | 79.3 | 34 | 43 | 25.2 | 42 | 21 | 0.0 |
| 10 | 14 | 179.8 | 23 | 13 | 78.9 | 35 | 15 | 25.1 | 42 | 38 | 0.0 |
| 11 | 12 | 163.2 | 24 | 32 | 76.9 | 36 | 48 | 21.9 | 42 | 39 | 0.0 |
| 12 | 3 | 152.6 | 25 | 49 | 61.7 | 37 | 4 | 7.5 | 42 | 46 | 0.0 |
| 13 | 40 | 144.3 | | | | | | | | | |

Figure 4-a illustrates the cumulative probability distribution of the BWC of the vertices. We notice that about 81% of the vertices have BWC values less than 180; while the largest BWC value observed is 821.4. Thus, the BWC metric exhibits a Power-law style distribution (Balakrishnan & Nevzorov, 2003) for the vertices in the US States graph. From Figure 4-b, we also notice that though vertices with a higher degree are more likely to have a higher BWC and the Spearman's rank-based correlation coefficient is 0.87; we do observe instances wherein the BWC values could vary appreciably even among vertices with the same degree centrality.



(a) Cumulative Probability Distribution of BWC (b) Degree Centrality vs. Betweenness Centrality

Figure 4. Betweenness Centrality of the Vertices in the US States Graph



(a) Cumulative Probability Distribution of ClC (b) Degree Centrality vs. Closeness Centrality

Figure 5. Closeness Centrality of the Vertices in the US States Graph

Table 5. Ranking of the Vertices in the US States Network Graph based on Closeness Centrality (ClC)

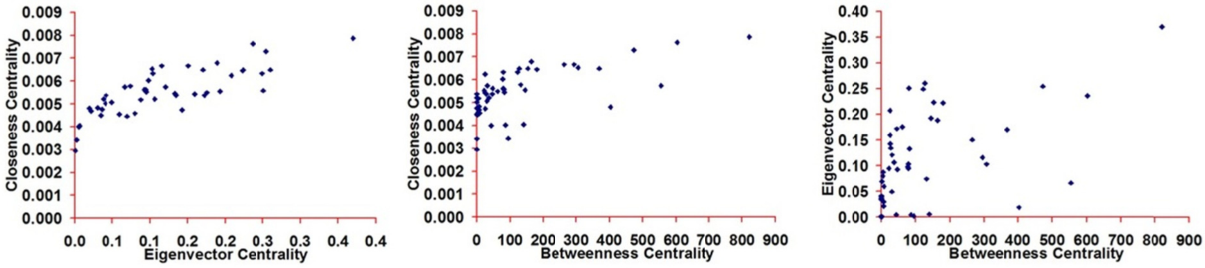
| Rank | ID | ClC | Rank | ID | ClC | Rank | ID | ClC | Rank | ID | ClC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 24 | 0.00787 | 10 | 15 | 0.00625 | 20 | 43 | 0.00543 | 29 | 2 | 0.00474 |
| 2 | 16 | 0.00763 | 11 | 32 | 0.00602 | 21 | 8 | 0.00538 | 30 | 29 | 0.00469 |
| 3 | 41 | 0.00730 | 12 | 19 | 0.00578 | 21 | 30 | 0.00538 | 31 | 25 | 0.00459 |
| 4 | 12 | 0.00680 | 13 | 23 | 0.00575 | 21 | 42 | 0.00538 | 32 | 4 | 0.00455 |
| 5 | 45 | 0.00667 | 13 | 37 | 0.00575 | 22 | 21 | 0.00521 | 33 | 9 | 0.00450 |
| 5 | 47 | 0.00667 | 14 | 1 | 0.00562 | 22 | 22 | 0.00521 | 34 | 33 | 0.00446 |
| 6 | 34 | 0.00654 | 14 | 10 | 0.00562 | 23 | 17 | 0.00518 | 35 | 20 | 0.00405 |
| 7 | 3 | 0.00649 | 15 | 5 | 0.00559 | 24 | 36 | 0.00508 | 36 | 44 | 0.00402 |
| 7 | 11 | 0.00649 | 16 | 40 | 0.00556 | 25 | 46 | 0.00503 | 37 | 6 | 0.00400 |
| 7 | 26 | 0.00649 | 17 | 48 | 0.00552 | 26 | 7 | 0.00483 | 38 | 28 | 0.00344 |
| 8 | 14 | 0.00645 | 18 | 49 | 0.00549 | 27 | 31 | 0.00481 | 39 | 38 | 0.00341 |
| 9 | 13 | 0.00633 | 19 | 27 | 0.00546 | 28 | 39 | 0.00476 | 40 | 18 | 0.00296 |
| 9 | 35 | 0.00633 | | | | | | | | | |

### *2.4 Closeness Centrality*

The Closeness Centrality (ClC) of a vertex is a measure of the number of hops on the shortest paths from the vertex to every other vertex in the graph. The ClC of a vertex is the inverse of the sum of the lengths (hops) of the shortest paths from the vertex to the rest of the vertices in the graph (determined using the Breadth First Search algorithm (Cormen et al., 2009)). We observe the state of Missouri to have the largest ClC, followed by the states of Kentucky and Tennessee. The state of Maine has the lowest ClC indicating that the sum of the length of the shortest paths from this state to the rest of the states is the largest. From Table 5, a total of 40 unique values could be observed for the ClC of the vertices (there are several instances where two or three states have the same values for the ClC). Figure 5-a captures the cumulative distribution of the ClC; we observe the values to be uniformly distributed albeit within a smaller range (unlike the EVC and BWC) resulting in a relatively steep curve. Figure 5-b captures the relationship between degree centrality and closeness centrality; vertices with a larger degree are likely to have a larger closeness centrality and the Spearman's rank-based correlation coefficient value is 0.75.

### *2.5 Ranking of Vertices Based on the Centrality Metrics*

Figures 3-b, 4-b and 5-b respectively capture the relationship between degree centrality and each of the other three centrality metrics (EVC, BWC and ClC). Figures 6-a, 6-b and 6-c capture the relationship between EVC, BWC and ClC. We present the values for the Spearman's rank-based correlation coefficient between any two centrality metrics in Table 6. The larger the value of the correlation coefficient for any two metrics, the more the similarity in the ranking of the vertices with respect to the two metrics in consideration. We observe the ranking between DegC and BWC to have the highest correlation coefficient (0.87), whereas the ranking between EVC and BWC has the lowest correlation coefficient (0.52). The lower correlation coefficient values for EVC-BWC and BWC-ClC metrics indicate the rankings of the vertices with respect to each of the two combinations are quite different from each other. Nevertheless, the state of Missouri has the largest value for all the four centrality metrics.



(a) EVC vs. ClC (b) BWC vs. ClC (c) BWC vs. EVC

Figure 6. Relationship between Eigenvector, Closeness and Betweenness Centrality Metrics for the US States Graph

Table 6. Spearman's Rank-based Correlation Coefficient among the Centrality Metrics for the US States Graph

| | EVC | BWC | ClC |
|---|---|---|---|
| DegC | 0.80 | 0.87 | 0.75 |
| EVC | | 0.52 | 0.80 |
| BWC | | | 0.68 |

### *2.6 Maximum and Maximal Clique Size*

A clique in a graph is a completely connected sub graph of the graph (Cormen et al., 2009); any two vertices within a clique are connected by an edge. The size of a clique is the number of vertices constituting the clique. The "maximum clique" is the clique of the largest size. However, not all vertices are likely to be part of the maximum clique (unless the graph is completely connected). In complex real-world networks, most of the vertices are likely to be part of cliques of size less than that of the maximum clique. The largest clique in which a vertex is part of is referred to as the maximal clique for the vertex and the corresponding clique size is referred to as the maximal clique size of the vertex. The problems of determining the maximum clique size for a graph as well as the maximal clique size for the individual vertices in a graph are NP-hard (Cormen et al., 2009) and one would need efficient heuristics to determine them. Pattabiraman et al (2016) proposed a branch-and-bound based heuristic to efficiently determine the maximum clique for complex network graphs and an extended version (Meghanathan, 2015b) of this heuristic could be used to determine the maximal clique size for the individual vertices in a graph. For the US States graph of 49 vertices, we observe the maximum clique size of the graph to be 4 comprising of the states of Arizona, Colorado, New Mexico and Utah (i.e., there is a shared boundary between any two of these four states), corresponding to the Four Corners Monument (Benson, 2008) and each of the remaining vertices (except Maine) have a maximal clique size of 3; the state of Maine has a maximal clique size of 2 as it has only one other state with which it shares a border. Table 7 lists the values for the centrality metrics for the four states constituting the maximum clique and the corresponding rank of these vertices with respect to these metrics is indicated in the parenthesis. We notice that the ranking of the vertices constituting the maximum clique is relatively higher with respect to the degree-based centrality metrics (DegC and EVC) compared to the shortest path-based centrality metrics (BWC and ClC). It appears that vertices constituting the maximum clique are not very critical to facilitate shortest path communication among the other vertices in the US States network graph.

Table 7. Centrality Metrics (Value and Rank) of the Vertices Constituting the Maximum Clique

| State (ID) | DegC (Rank) | EVC (Rank) | BWC (Rank) | ClC (Rank) |
|---|---|---|---|---|
| Arizona (2) | 5 (5) | 0.1423 (17) | 25.4 (33) | 0.00474 (29) |
| Colorado (5) | 7 (3) | 0.2509 (4) | 80.4 (21) | 0.00559 (15) |
| New Mexico (30) | 5 (5) | 0.1720 (13) | 45.5 (27) | 0.00538 (21) |
| Utah (43) | 5 (5) | 0.1595 (15) | 25.2 (34) | 0.00543 (20) |

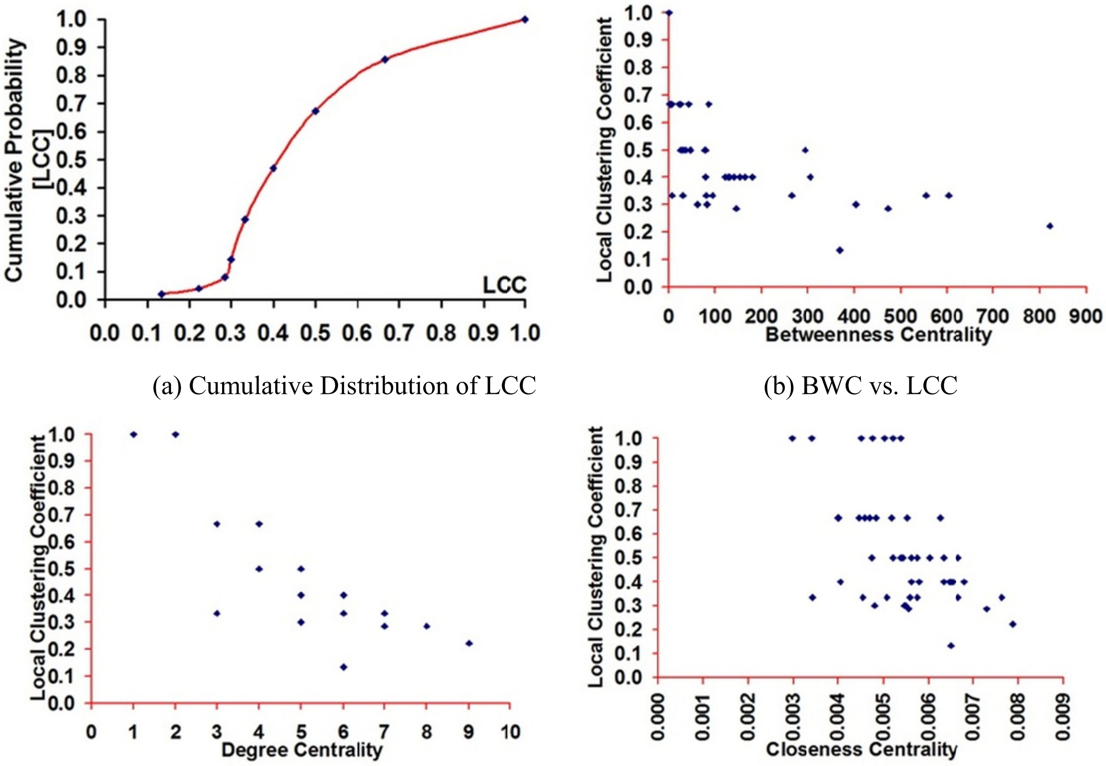
### *2.7 Local Clustering Coefficient*

The local clustering coefficient (LCC) of a vertex is a measure of the probability that any two neighbors of the vertex are connected. For a vertex vi with ki neighbors, the maximum number of links between any two neighbors of the vertex is ki(ki-1)/2. The LCC of a vertex is the ratio of the actual number of links connecting the neighbors of the vertex to that of the maximum possible number of links between the neighbors of the vertex. The smaller the LCC of a vertex, the more important is the vertex for facilitating shortest path communication among its neighbors (as there is a good chance that the neighbors of a vertex that are connected to each other go through the vertex for shortest path communication). Hence, we give a higher rank to vertices having a lower LCC.

Table 8. Ranking of the Vertices in the US States Network Graph based on Local Clustering Coefficient (LCC)

| Rank | ID | LCC | Rank | ID | LCC | Rank | ID | LCC | Rank | ID | LCC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 11 | 0.133 | 5 | 45 | 0.333 | 7 | 13 | 0.500 | 8 | 25 | 0.667 |
| 2 | 24 | 0.222 | 6 | 3 | 0.400 | 7 | 22 | 0.500 | 8 | 29 | 0.667 |
| 3 | 40 | 0.286 | 6 | 10 | 0.400 | 7 | 23 | 0.500 | 8 | 33 | 0.667 |
| 3 | 41 | 0.286 | 6 | 12 | 0.400 | 7 | 30 | 0.500 | 8 | 44 | 0.667 |
| 4 | 27 | 0.300 | 6 | 14 | 0.400 | 7 | 32 | 0.500 | 8 | 48 | 0.667 |
| 4 | 31 | 0.300 | 6 | 19 | 0.400 | 7 | 42 | 0.500 | 9 | 8 | 1.000 |
| 4 | 49 | 0.300 | 6 | 20 | 0.400 | 7 | 43 | 0.500 | 9 | 9 | 1.000 |
| 5 | 4 | 0.333 | 6 | 26 | 0.400 | 7 | 47 | 0.500 | 9 | 18 | 1.000 |
| 5 | 5 | 0.333 | 6 | 34 | 0.400 | 8 | 6 | 0.667 | 9 | 21 | 1.000 |
| 5 | 16 | 0.333 | 6 | 35 | 0.400 | 8 | 7 | 0.667 | 9 | 38 | 1.000 |
| 5 | 28 | 0.333 | 7 | 1 | 0.500 | 8 | 15 | 0.667 | 9 | 39 | 1.000 |
| 5 | 36 | 0.333 | 7 | 2 | 0.500 | 8 | 17 | 0.667 | 9 | 46 | 1.000 |
| 5 | 37 | 0.333 | | | | | | | | | |

Table 8 ranks the vertices in the US States graph in the increasing order of the values of the LCC. As the LCC values get larger, we observe a significant number of ties among the vertices. The state of Idaho (with a degree of 6) has the lowest LCC and hence is the top ranked with respect to the LCC metric. The state of Missouri (that was ranked first with respect to all the four centrality metrics) is ranked second with respect to LCC. There are only nine unique values for the LCC metric. Figure 7-a captures the cumulative probability distribution of the LCC metric and we observe that only about 15% of the vertices have a LCC of 0.3 or lower, and more than half of these vertices have the largest values for the BWC (as observed in Figure 7-b). We observe the Spearman's Rank-based correlation coefficient between LCC and BWC (computed based on the rankings in Tables 4 and 8) to be 0.82. Figure 7-c very well captures the inverse relationship between degree and LCC. Vertices having a larger degree are more likely to have a lower LCC as it would be difficult to expect any two neighbors of a high-degree node to be directly connected to each other and are more likely to go through the vertex for shortest-path communication. On the other hand, vertices having a lower degree are more likely to have a larger LCC as it is highly possible for any two neighbors of a low-degree vertex to be directly connected to each other and need not go through the vertex for shortest path communication. Thus, vertices with higher degree and lower LCC are more likely to have a larger BWC, and vertices with a lower degree and higher LCC are more likely to have a smaller BWC. A plot of Closeness Centrality (ClC) vs. LCC reveals that the two metrics are almost independent of each other (as vertices covering the entire range of values observed for the ClC have almost the same LCC), leading to a Spearman's rank-based correlation coefficient of 0.52.



(a) Cumulative Distribution of LCC (b) BWC vs. LCC

(c) DegC vs. LCC (d) ClC vs. LCC

Figure 7. Cumulative Distribution of the Local Clustering Coefficient and its Relationship with the Degree Centrality, Betweenness Centrality and Closeness Centrality

### *2.8 Distance Metrics*

We determine the distribution of the path length for any two vertices and eccentricity (Ecc) of the vertices in the US States network graph. The path length for a pair of vertices (*u*, *v*) is the number of hops in the shortest path (minimum hop path) between the two vertices. The eccentricity of a vertex is the maximum of the path lengths to any other vertex in the graph (Newman, 2010). In other words, the eccentricity of a vertex is the maximum of the minimum number of hops to any other vertex. The eccentricity of a particular vertex and the path lengths of the vertex to every other vertex are determined by running the Breadth First Search algorithm (Cormen et al., 2009) on the vertex. Figure 8 illustrates the distribution of the path length of the vertex pairs and eccentricity of the vertices in the US States graph. We observe the path length distribution to be Poisson in nature with a long tail (Kurtosis of 2.82) and mean value of 3.93 as well as a standard deviation of 1.98. The average path length value of 3.93 is close to the expected average path value (of 3.89) to exhibit the small-world property (path length of ln(*n*) for a graph of *n* vertices; Newman, 2010) for a graph of 49 vertices.

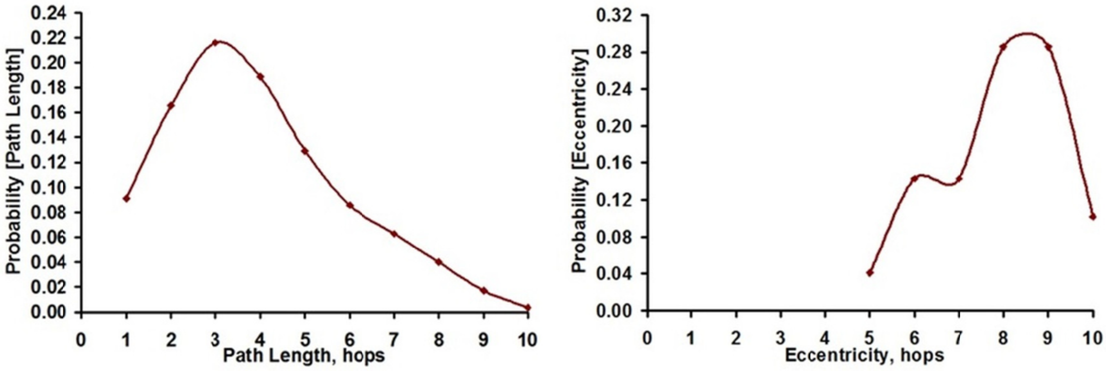


Figure 8. Distribution of Path Length and Eccentricity for the US States Graph

Table 9. Ranking of the Vertices in the US States Network Graph based on Eccentricity (Ecc)

| Rank | ID | Ecc | Rank | ID | Ecc | Rank | ID | Ecc | Rank | ID | Ecc |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 34 | 5 | 3 | 31 | 7 | 4 | 35 | 8 | 5 | 42 | 9 |
| 1 | 47 | 5 | 3 | 32 | 7 | 4 | 39 | 8 | 5 | 43 | 9 |
| 2 | 8 | 6 | 3 | 41 | 7 | 4 | 44 | 8 | 5 | 46 | 9 |
| 2 | 13 | 6 | 4 | 1 | 8 | 4 | 48 | 8 | 5 | 49 | 9 |
| 2 | 16 | 6 | 4 | 3 | 8 | 5 | 5 | 9 | 6 | 2 | 10 |
| 2 | 19 | 6 | 4 | 6 | 8 | 5 | 9 | 9 | 6 | 4 | 10 |
| 2 | 21 | 6 | 4 | 10 | 8 | 5 | 17 | 9 | 6 | 18 | 10 |
| 2 | 37 | 6 | 4 | 11 | 8 | 5 | 22 | 9 | 6 | 25 | 10 |
| 2 | 45 | 6 | 4 | 14 | 8 | 5 | 27 | 9 | 6 | 33 | 10 |
| 3 | 7 | 7 | 4 | 15 | 8 | 5 | 28 | 9 | | | |
| 3 | 12 | 7 | 4 | 20 | 8 | 5 | 30 | 9 | | | |
| 3 | 24 | 7 | 4 | 23 | 8 | 5 | 36 | 9 | | | |
| 3 | 29 | 7 | 4 | 26 | 8 | 5 | 38 | 9 | | | |

The distribution of the eccentricity of the vertices shows that the minimum value (also called radius): 5 is half of the maximum value (also called diameter): 10. Nevertheless, we observe that more than 65% of the vertices have an eccentricity of 8 or above (i.e., more than 65% of the vertices have a maximum path length of 8-10 to one or more vertices) and only 4% of the 49 vertices (i.e., just 2 vertices) incur eccentricity values corresponding to the radius of the graph. The two states of West Virginia and Ohio (with an eccentricity corresponding to the radius) are said to form the "center" of the graph (Newman, 2010); each of these two vertices are within a maximum hop count of 5 on a shortest path to any other vertex in the graph. Note that neither of these two vertices are among the vertices that are ranked in the top 3 with respect to any of the centrality metrics and local clustering coefficient. There are five states (Arizona, California, Maine, Montana and North Dakota) that have an eccentricity corresponding to the diameter of the graph. Table 9 illustrates a ranking of the vertices based on eccentricity (the state with the smallest eccentricity is ranked first).

## 3. Network-Level Metrics

In this section, we evaluate the following network-level metrics for the US States graph: Bipartivity Index; Degree Metrics - Average, Standard Deviation, Kurtosis and Spectral Radius Ratio; Algebraic Connectivity; Assortativity Index and Modularity. We also determine the size of the Minimum Connected Dominating Set of vertices based on the four centrality metrics (DegC, BWC, EVC and ClC).

### *3.1 Bipartivity Index*

A graph is bipartite (a.k.a. 2-colorable) if the vertices of the graph can be partitioned to two disjoint sets such that all the edges in the graph are those that connect a vertex from one partition to the other partition, and there are no edges between vertices within a partition (Cormen et al., 2009). The two partitions are determined using the sign of the entries in the eigenvector corresponding to the smallest eigenvalue of the binary adjacency matrix of the graph (Estrada & Rodriguez-Velazquez, 2005); the positive entries are grouped into one partition and the negative entries are grouped into another partition. Figure 9 displays the US States graph with the states colored in yellow or green to represent the two partitions.

A measure called bipartivity index (Estrada & Rodriguez-Velazquez, 2005) has been proposed in the literature to determine the extent of bipartivity for complex network graphs. The bipartivity index of a graph is computed using the eigenvalues of the binary adjacency matrix of the graph. The bipartivity index values could range from 0 to 1; if a graph has bipartivity index of 1, it implies all the edges in the graph are only those that connect the vertices across the two partitions. However, there exist several real-world network graphs for which there are few edges (called frustrated edges) that connect the vertices within each partition (though a majority of the edges connect the vertices across two partitions; Estrada & Rodriguez-Velazquez, 2005). Graphs with one or more frustrated edges have bipartivity index less than 1 and graphs with no frustrated edges have bipartivity index equal to 1 (Estrada & Rodriguez-Velazquez, 2005). While graphs with no frustrated edges have been referred to as *truly bipartite*, graphs with frustrated edges have been referred to as *close-to-bipartite* (Estrada & Rodriguez-Velazquez, 2005). The bipartivity index of the US States graph has been observed to be 0.66 and the fraction of frustrated edges in the network is 0.32. Though the bipartivity index value is not that close to 1, it is still larger than the values observed for several of the real-world networks in the literature (Estrada & Rodriguez-Velazquez, 2005).

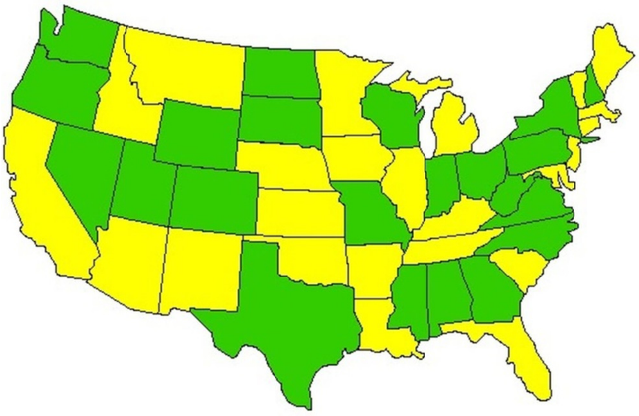


Figure 9. On the Bipartivity of the US States Network Graph (48 Contiguous States and DC)

Bipartivity Index: 0.66

### 3.2 Degree Metrics

From Figure 2, we observe that the degree distribution of the vertices is Poisson and bi-modal in nature. To corroborate this assertion, we observe the average degree of the vertices to be 4.37 (roughly close to the average of the two peak degree values of 3 and 5), with a standard deviation and Kurtosis of 1.72 and 2.75 respectively (all indicating that the degree distribution is close to being a normal/Poisson distribution; Balakrishnan & Nevzorov, 2003). Kurtosis is a measure of the "tailedness" of a probability distribution (Balanda & MacGillivray, 1998). For normal distribution, the expected value for the Kurtosis is 3; distributions with Kurtosis values above or below 3 are said to be fat-tailed and thin-tailed respectively (Balanda & MacGillivray, 1998). Further, the spectral radius ratio for node degree (defined as the ratio of the principal eigenvalue of the adjacency matrix of the graph and the average node degree; Meghanathan, 2014a) is observed to be 1.24. The spectral radius ratio for node degree is a measure of the variation in node degree of the vertices; the minimum possible value for this metric is 1.0 and farther the value of the spectral radius ratio for node degree from 1, the larger is the variation in node degree. Though the spectral radius ratio for node degree value of 1.24 is not much larger than 1, the value is not as close to 1 (Meghanathan, 2014a) as observed for the US Football network (Girvan & Newman, 2002): a network with a unimodal Poisson degree distribution of the vertices; the value of 1.24 could be attributed to the variation due to the bi-modal degree distribution of the vertices.

### 3.3 Algebraic Connectivity

The algebraic connectivity of a graph is a measure of the robustness of the graph (Fiedler, 1973). The farther the value of this metric from 0, the larger is the robustness of the graph with respect to the overall connectivity of the network. The algebraic connectivity of a graph is computed as the second smallest eigenvalue of the Laplacian matrix of the graph [24]. We determine the algebraic connectivity of the US States graph to be 0.0973. Such a low value indicates that (though the entire graph is connected), the robustness of the graph is very low. The entries in the Laplacian matrix of a graph are defined (Fiedler, 1973) as follows:

$$L(i,j) = \begin{Bmatrix} degree(i) & if\ i = j \\ -\ a_{ij} & if\ i \neq j \end{Bmatrix}$$ where $a_{ij}$ is an entry (0 or 1) in the adjacency matrix for vertices $i$ and $j$.

### 3.4 Assortativity

The assortative index of the edges (with respect to a particular node-level metric like centrality metrics) in a graph is a measure of the extent of similarity (with respect to the metric) between the end vertices of the graph (Newman, 2010). Assortative index with respect to a particular metric is computed as the Pearson's product-moment correlation coefficient values (with respect to the metric in consideration) for the end vertices of the edges in the graph. Like correlation coefficient, the values for the Assortative index could range from -1 to 1 (Newman, 2010). If the Assortative index value is close to 1 (or -1), it implies the end vertices of the edges in the graph have similar values (or dissimilar values) for the particular metric in consideration. Networks with assortative index value closer to 1 are said to be more assortative and value closer to -1 are said to be more dissortative with respect to the node-level metric in consideration (Meghanathan, 2016a). If the Assortative index value is closer to 0, it implies the values for the end vertices of the edges are independent of each other with respect to the particular metric in consideration. Random networks are expected to have an Assortative index closer to 0 for any node-level metric (Meghanathan, 2016a).

We conduct an assortativity analysis of the edges in the US States graph with respect to the four centrality metrics (DegC, EVC, BWC and ClC) and local clustering coefficient (LCC) and observe the following values: (i) DegC-based Assortativity index: 0.23; (ii) EVC-based Assortativity index: 0.62; (iii) BWC-based Assortativity index: 0.23; (iv) ClC-based Assortativity index: 0.65 and (v) LCC-based Assortativity index: -0.03. We thus observe the US States graph to be relatively more assortative with respect to EVC and ClC and less assortative with respect to DegC and BWC; also, the US States graph is neither assortative nor dissortative with respect to LCC.

Table 10. Partitioning of the Vertices of the Contiguous States Graph (48 States and DC) into Communities (using the Louvain Algorithm; Blondel et al., 2008)

| 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | State | ID | State | ID | State | ID | State | ID | State | ID | State |
| 1 | AL | 2 | AZ | 25 | MT | 6 | CT | 7 | DE | 12 | IL |
| 3 | AR | 4 | CA | 26 | NE | 18 | ME | 8 | DC | 13 | IN |
| 9 | FL | 5 | CO | 33 | ND | 20 | MA | 19 | MD | 14 | IA |
| 10 | GA | 11 | ID | 40 | SD | 28 | NH | 29 | NJ | 16 | KY |
| 17 | LA | 15 | KS | 49 | WY | 31 | NY | 37 | PA | 21 | MI |
| 23 | MS | 27 | NV | | | 38 | RI | 45 | VA | 22 | MN |
| 32 | NC | 30 | NM | | | 44 | VT | 47 | WV | 24 | MO |
| 39 | SC | 35 | OK | | | | | | | 34 | OH |
| 41 | TN | 36 | OR | | | | | | | 48 | WI |
| 42 | TX | 43 | UT | | | | | | | | |
| | | 46 | WA | | | | | | | | |

### *3.5 Modularity*

We used the Louvain algorithm (Blondel et al., 2008) to determine an optimal partitioning of the US States graph into communities. The modularity score (in a scale of 0 to 1; Newman, 2006) was observed to be 0.586 and there were a total of 6 communities of the vertices (see Table 10 and Figure 10). We observe the vertex communities to closely resemble the nine regional divisions used by the United States Census Bureau (http://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf). Figure 10 displays the six communities with different colors (one color per community) using the map from http://www.thecolor.com.

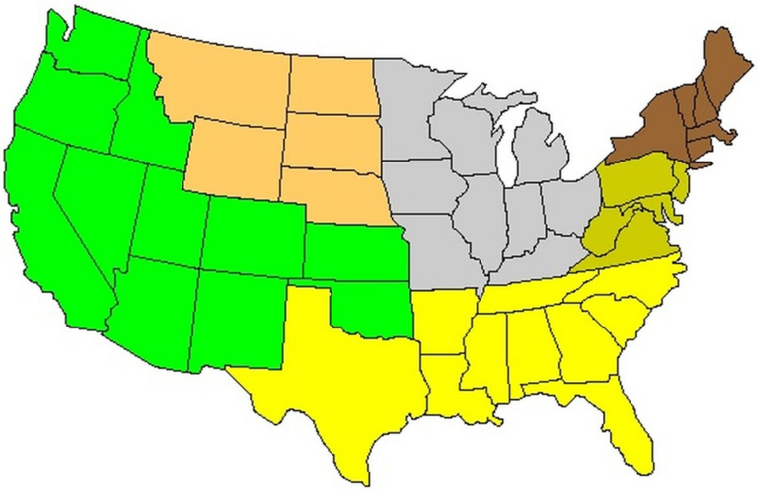
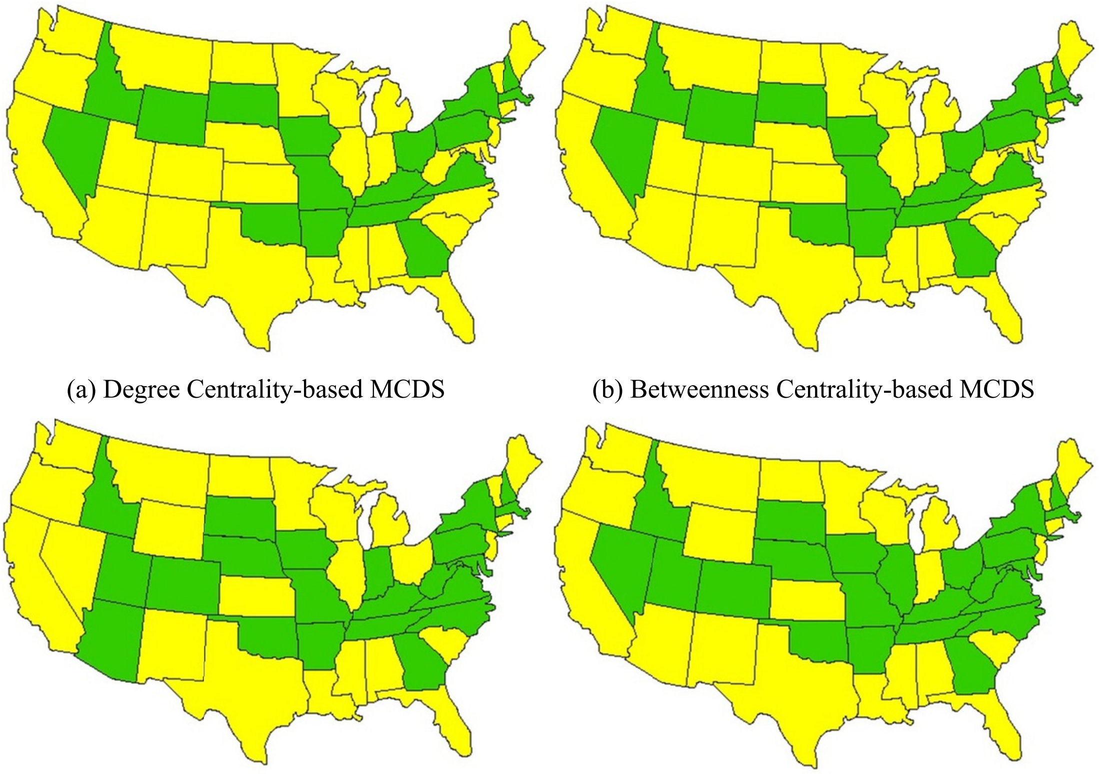


Figure 10. Communities of Vertices in the US States Graph (48 Contiguous States and DC) Detected using the Louvain Algorithm (Modularity Score: 0.586)



(a) Degree Centrality-based MCDS

(b) Betweenness Centrality-based MCDS

(c) Eigenvector Centrality-based MCDS

(d) Closeness Centrality-based MCDS

Figure 11. Approximations to the Minimum Connected Dominating Set (MCDS) based on Centrality Metrics for the US States Graph (48 Contiguous States and DC)

### *3.6 Connected Dominating Set*

A connected dominating set (CDS) of a graph is a subset of the vertices such that every vertex in the graph is either in the CDS or is a neighbor of a node in the CDS (Cormen et al., 2009). The problem of determining a minimum connected dominating set (MCDS) is NP-hard (Cormen et al., 2009) and there are several heuristics (e.g., Newman, 2006; Meghanathan, 2014b) available in the literature to approximate a MCDS. We use the heuristic proposed by Meghanathan (2014b) to determine an approximate MCDS with respect to a chosen node-level metric. The idea behind the MCDS-heuristic is to prefer to include nodes with the largest value for the node-level metric in consideration to be part of the CDS; once a node is included is to the CDS, all its neighbor nodes are said to be covered and are considered for possible inclusion to the CDS. The heuristic proceeds in iterations; in each iteration, a covered node with the largest value for the node-level metric is included to the CDS as long as the covered node has at least one uncovered neighbor node that is not yet covered by nodes already included to the CDS. The iterations are continued until all nodes are covered (i.e., a node is either in the CDS or is covered by a node in the CDS).

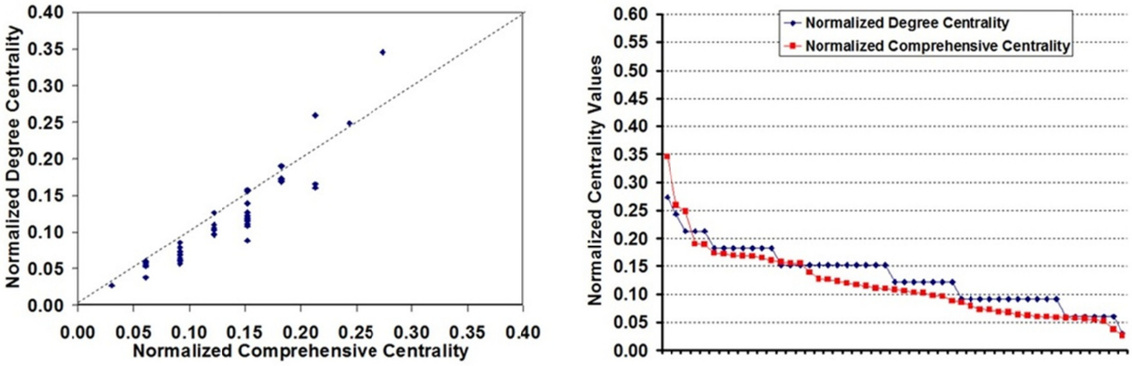
Any node-level metric could be used to approximate a MCDS; however, the size of the MCDS varies with the node-level metric used. Traditionally, the degree centrality metric has been observed to return CDSs of the smallest size (Newman, 2006; Meghanathan, 2014b) as a high-degree node included to the CDS is more likely to cover several other nodes. In this paper, we indeed observe the above assertion to be true as a degree-based MCDS of the US States graph of 49 vertices has only 17 vertices; we also observe the BWC-MCDS of the US States graph to comprise of the same minimum number of vertices (i.e., 17 vertices) and the constituent states for both the MCDSs are the same. We observe the EVC and ClC-based MCDSs to incur a relatively larger number of vertices (22 and 23 vertices respectively). Figures 11-(a) through 11-(d) present the MCDSs of the US States network graph (vertices that are part of the MCDS are colored in green and vertices that are not part of the MCDS, but covered by at least one node in the MCDS are colored in yellow).

## 4. Normalization-based Comprehensive Centrality Scores

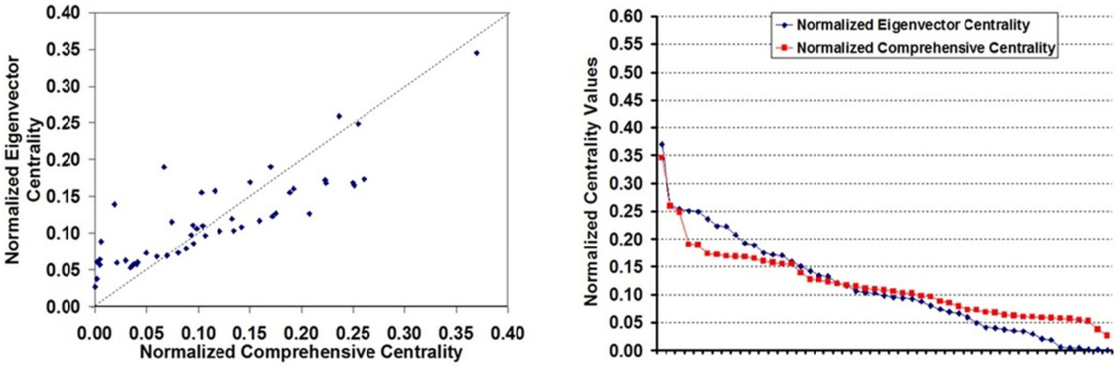
As there are ties among the vertices when ranked with respect to a particular centrality metric, we propose a normalization-based approach to obtain a potentially tie-free comprehensive ranking of the vertices in the US states network graph. In addition to the above objective, we seek to identify the centrality metric whose normalized values are relatively the closest to the NCC values as well as we intend to identify the centrality metric whose ranking of the vertices matches relatively closest to the ranking based on the NCC values. If we could identify such centrality metrics with a lower root mean square difference (RMSD) value, we could just compute these centrality metric(s) and consider the ranking based on these metric(s) as a comprehensive measure of ranking the vertices rather than individually computing the various centrality metrics. A similar approach could be applied for any real-world network or synthetic networks generated from theoretical models (Erdos & Renyi, 1959; Barabasi & Albert, 1999).

Table 11. Normalized Comprehensive Centrality (NCC)-based Ranking of Vertices in the US States Graph

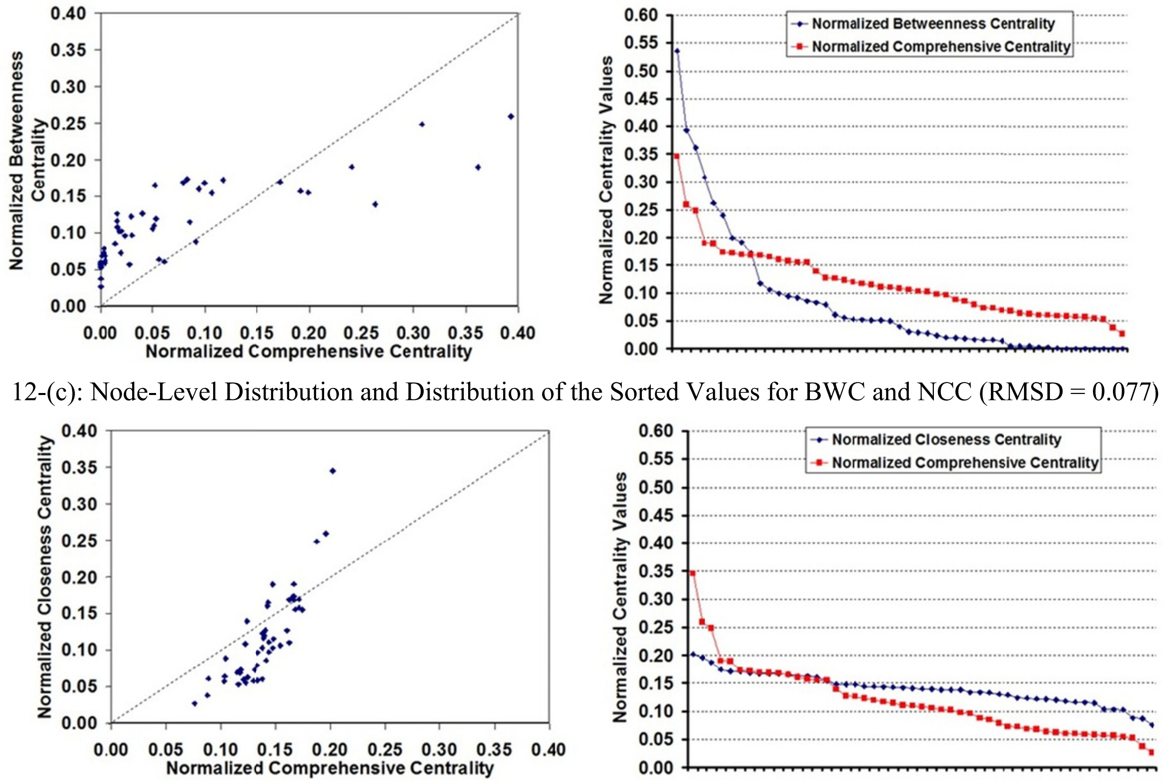
| Rank | ID | NCC | Rank | ID | NCC | Rank | ID | NCC | Rank | ID | NCC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 24 | 0.3454 | 14 | 34 | 0.1556 | 26 | 32 | 0.1062 | 38 | 44 | 0.0639 |
| 2 | 16 | 0.2595 | 15 | 12 | 0.1555 | 27 | 42 | 0.1032 | 39 | 7 | 0.0624 |
| 3 | 41 | 0.2485 | 16 | 31 | 0.1394 | 28 | 23 | 0.1027 | 40 | 28 | 0.0608 |
| 4 | 11 | 0.1900 | 17 | 49 | 0.1272 | 29 | 1 | 0.0974 | 41 | 8 | 0.0601 |
| 5 | 37 | 0.1896 | 18 | 15 | 0.1266 | 30 | 22 | 0.0965 | 42 | 29 | 0.0594 |
| 6 | 26 | 0.1732 | 19 | 30 | 0.1230 | 31 | 20 | 0.0884 | 43 | 21 | 0.0582 |
| 7 | 14 | 0.1721 | 20 | 27 | 0.1197 | 32 | 48 | 0.0856 | 44 | 46 | 0.0576 |
| 8 | 45 | 0.1692 | 21 | 43 | 0.1169 | 33 | 17 | 0.0789 | 45 | 6 | 0.0568 |
| 9 | 35 | 0.1684 | 22 | 19 | 0.1151 | 34 | 25 | 0.0731 | 46 | 39 | 0.0546 |
| 10 | 3 | 0.1681 | 23 | 10 | 0.1107 | 35 | 36 | 0.0728 | 47 | 9 | 0.0527 |
| 11 | 5 | 0.1649 | 24 | 13 | 0.1099 | 36 | 33 | 0.0691 | 48 | 38 | 0.0376 |
| 12 | 40 | 0.1606 | 25 | 2 | 0.1082 | 37 | 4 | 0.0682 | 49 | 18 | 0.0267 |
| 13 | 47 | 0.1578 | | | | | | | | | |



12-(a): Node-Level Distribution and Distribution of the Sorted Values for Degree and NCC (RMSD = 0.027)



12-(b): Node-Level Distribution and Distribution of the Sorted Values for EVC and NCC (RMSD = 0.047)



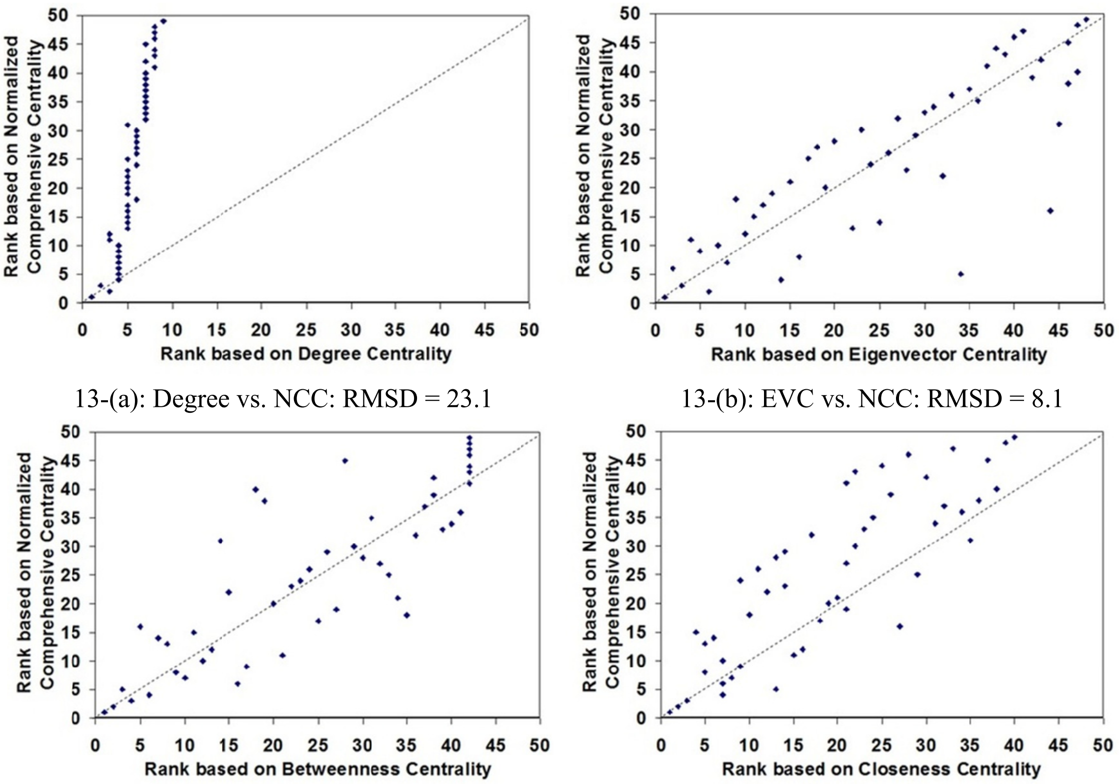
12-(c): Node-Level Distribution and Distribution of the Sorted Values for BWC and NCC (RMSD = 0.077)

12-(d): Node-Level Distribution and Distribution of the Sorted Values for ClC and NCC (RMSD = 0.047)

Figure 12. Node-Level Distribution of the Normalized Comprehensive Centrality Scores and the Normalized Centrality Scores for the Individual Metrics: (a) Degree, (b) EVC, (c) BWC and (d) ClC

We normalize the centrality values for each of the four metrics: degree, eigenvector, betweenness and closeness, and compute a normalized comprehensive centrality (NCC) of the vertices as a weighted average of the normalized values of the centrality metrics. There could be a tie between two or more vertices with respect to the NCC values if and only if the corresponding vertices incur identical values for the four centrality metrics that are part of the NCC formulation. In this paper, we assign equal weights (0.25 each) for the four centrality metrics that are used to compute the NCC values. In general, this idea could be used to compute a normalized comprehensive centrality score involving any number of centrality metrics and with different weight for each metric as long as the sum of the weights is 1.0. Since larger the value for an individual centrality metric (for all the four centrality metrics that we use to compute the NCC), the higher is the ranking of the vertex with respect to the centrality metric, we propose that the larger the NCC value for a vertex, the higher is the overall ranking of the vertex. One can observe in Table 11: the NCC values are unique for the vertices and the ranking of the vertices could be done without any ties. ThisWe observe the states of Missouri, Kentucky and Tennessee to obtain the top three ranking with respect to NCC. At least two of these three states obtain the top three ranking with respect to each of the four centrality metrics. We observe the state of Maine to obtain the bottommost ranking with respect to the NCC scores and this state also obtained the bottommost ranking for three of the above four centrality metrics.

In Figures 12-(a) through 12-(d), we compare the normalized comprehensive centrality values to that of the individual normalized centrality values as well as illustrate the distribution of the normalized values for each of the four centrality metrics vis-a-vis the normalized comprehensive centrality value (each shown in the order of the largest to the smallest). For a majority of the vertices: we observe the normalized comprehensive centrality values to be lower than that of normalized degree (see Figure 12-a) and closeness centrality metric values (see Figure 12-d); whereas, we observe the normalized comprehensive centrality values to be larger than that of the normalized betweenness centrality values (see Figure 12-b). On the other hand, we observe the data points for the normalized comprehensive centrality and the normalized eigenvector centrality to be evenly distributed above and below the diagonal line (see Figure 12-b).



13-(a): Degree vs. NCC: RMSD = 23.1

13-(b): EVC vs. NCC: RMSD = 8.1

13-(c): BWC vs. NCC: RMSD = 7.7

13-(d): ClC vs. NCC: RMSD = 9.6

Figure 13. Distribution of the Ranking of the Vertices based on the Normalized Comprehensive Centrality Scores and the Normalized Centrality Scores for the Individual Metrics: (a) Degree, (b) EVC, (c) BWC and (d) ClC

We computed the root mean squared difference (RMSD) between the NCC values and the values for each of the four centrality metrics. The RMSD values are shown along with the charts in Figures 12-(a) through 12-(d). We observe the degree centrality-NCC combo to have relatively the lowest RMSD value (0.027), while the BWC-NCC combo incur relatively the largest RMSD value (0.077). The EVC-NCC combo and ClC-NCC combo incur RMSD value of 0.047 each. Thus, for the US States network graph, the degree centrality appears to be the centrality metric whose normalized values are relatively the closest to the NCC values.

Similar to the approach taken above, we identify the centrality metric whose ranking for the vertices matches relatively the closest to the ranking of the vertices based on NCC. In Figures 13-(a) through 13-(d), we show plots of the numerical ranking of the vertices based on each of the four centrality metrics vis-a-vis the NCC. For all the four centrality metrics, we notice vertices that were ranked high (i.e., lower numerical value for the rank) are also more likely to receive a higher rank with respect to NCC. We computed the RMSD values for the ranking obtained with NCC and each of the four centrality metrics. We observe the BWC-NCC combo to incur the lowest RMSD value of 7.7, closely followed by the EVC-NCC combo (with a RMSD value of 8.1). The degree centrality-NCC combo incurred the largest RMSD value of 23.1, primarily attributed to the narrow range of values [1...9] for node degree and the broader range of values [1...49] for the NCC.

Table 12 summarizes the RMSD values obtained for the ranking of the vertices based on the normalized individual centrality scores and the normalized comprehensive centrality scores (Figures 13-a through 13-d) and the actual scores themselves (Figures 12-a through 12-d). The degree cenntrality metric incurs the lowest RMSD value based on the normalized centrality scores and the largest RMSD value based on the ranking of the vertices. On the other hand, the betweenness centrality metric incurs the lowest RMSD value based on the ranking of the vertices and the largest RMSD value based on the normalized centrality scores. The closeness centrality metric incurs the second largest RMSD value based on the ranking of the vertices. Considering all of the above, the eigenvector centrality metric (EVC), which incurs the second lowest RMSD value with respect to both the normalized centrality scores and the ranking of the vertices, could be claimed as the centrality metric that closely matches to the normalized comprehensive centrality (NCC) values computed for the vertices of the US States network graph.

Table 12. Root Mean Square Difference (RMSD) Values obtained for the Node-Level Distribution of the Normalized Centrality Scores and the Ranking of the Vertices based on the Normalized Scores vis-a-vis the Normalized Comprehensive Centrality (NCC) Scores

| Centrality Metric-NCC Combo | Node-Level Distribution of the Normalized Values | Ranking of the Vertices based on the Normalized Values |
|---|---|---|
| Degree-NCC | 0.027 | 23.1 |
| Eigenvector-NCC | 0.047 | 8.1 |
| Betweenness-NCC | 0.077 | 7.7 |
| Closeness-NCC | 0.047 | 9.6 |

## 5. Configuration Model-Based Analysis

Given the degree sequence of a real-world network, the Configuration model could be used to generate a random network whose degree sequence is also the same as that of the real-world network (i.e., the random network could even have a non-Poisson degree distribution if the corresponding real-world network has one; Meghanathan, 2016c). In this paper, we use the Configuration model to study whether the degree sequence of the US States network graph (a real-world network) would be sufficient to generate a random network whose node-level metrics and network-level metrics exhibit strong correlation or proximity with the values incurred for these metrics in the corresponding real-world network.

Let *N* and *L* be respectively the number of nodes and edges in the chosen real-world network of study (like the US States network graph). Given the degree sequence (D) for the chosen real-world network, we simulate the generation of a random network per the configuration model as follows: We create a list LD (of length corresponding to the sum of the node degrees): the list is initialized with node IDs and the number of instances a node ID appears in the list corresponds to the degree of the node in D. The list LD is shuffled. We then proceed in iterations (to generate the random network), traversing the list LD in the reverse direction (i.e., with index *j* from |LD| to 2). In each iteration: we generate an edge (for the random network) involving the vertex at index *j* in the list LD to a vertex at a randomly chosen index *i* (*i* < *j*) when the following conditions are met: (i) the two entries are not -1, (ii) the two vertices are not the same (to avoid self-loop) and (iii) there does not exist already an edge involving the two vertices in the random network. The entries at both the indexes *i* and *j* are then set to -1.

Table 13. Correlation of the Node-Level Metrics for the US States Network Graph and its 100 Instances of Random Networks (with the same Degree Sequence) Generated using the Configuration Model

| Node-Level Metric | Correlation Coefficient Value | Level of Correlation |
|---|---|---|
| Degree Centrality | 0.99 | Very Strongly Positive |
| Closeness Centrality | 0.99 | Very Strongly Positive |
| Eigenvector Centrality | 0.76 | Strongly Positive |
| Betweenness Centrality | 0.72 | Strongly Positive |
| Eccentricity | 0.48 | Moderately Positive |
| Maximal Clique Size | 0.38 | Weakly Positive |
| Local Clustering Coefficient | 0.33 | Weakly Positive |

Table 14. Correlation of the Node-Level Metrics for the US States Network Graph and its 100 Instances of Random Networks (with the same Degree Sequence) Generated using the Configuration Model

| Network-Level Metric | Real-World Network (US States Network Graph) | Random Network (Configuration Model) |
|---|---|---|
| Average Path Length | 3.93 | 2.61 |
| Diameter | 10.00 | 5.28 |
| Bipartivity Index | 0.66 | 0.85 |
| Algebraic Connectivity | 0.097 | 0.645 |
| Spectral Radius | 1.24 | 1.19 |
| Modularity Score | 0.58 | 0.88 |
| Edge Assortativity (Degree) | 0.23 | 0.23 |
| Edge Assortativity (EVC) | 0.63 | 0.17 |
| Edge Assortativity (BWC) | 0.23 | 0.16 |
| Edge Assortativity (ClC) | 0.65 | 0.18 |

We generate 100 instances of random networks for the US States network graph according to the Configuration model and measure the following node-level metrics: (i) Degree Centrality, (ii) Eigenvector Centrality, (iii) Betweenness Centrality, (iv) Closeness Centrality, (v) Maximal Clique Size, (vi) Local Clustering Coefficient and (vii) Eccentricity; and network-level metrics: (i) Assortativity Index of the edges based on each of the four centrality metrics, (ii) Spectral Radius Ratio for Node Degree, (iii) Average Path Length, (iv) Diameter, (v) Bipartivity Index, (vi) Algebraic Connectivity and (vii) Modularity score determined using the Louvain algorithm. In the case of the node-level metrics, we measured the Pearson's product-moment correlation coefficient (Triola, 2012) between the values incurred for the nodes in each of the 100 instances of the random networks and the actual real-world network and averaged the correlation coefficient values (shown in Table 13 in the decreasing order of the correlation coefficient values).

We adapt the range of correlation coefficient values (rounded to two decimals) proposed in the literature (Evans, 1995) to decide on the level of correlation. We observe a very strong positive correlation (range: 0.80...1.00) in the case of the degree centrality (as expected) and closeness centrality metrics, and a strongly positive correlation (range: 0.60...0.79) in the case of the eigenvector centrality and betweenness centrality metrics. On the other hand, we observe a moderately positive correlation (range: 0.40...0.59) in the case of eccentricity, and a weakly positive correlation (range: 0.20...0.39) in the case of maximal clique size and local clustering coefficient.

For each network-level metric, we averaged the results obtained with the 100 instances of the random networks and compared this average value with the value incurred for the actual US States network graph (shown in Table 14). For none of the network-level metrics (other than degree-based edge assortativity and spectral radius ratio for node degree), we observe the average values obtained for the random networks generated using the configuration model to be closer to the values obtained for the actual US States network graph. We observe the random network instances to be relatively more bipartite, more robust to disconnection and more modular. We also observe the random network instances to have a relatively smaller diameter and a smaller average path length between any two nodes. As expected of a random network, we also observe the edges to be very weakly assortative with respect to all the four centrality metrics for the random networks generated using the configuration model; on the other hand, we observe the edges to be strongly assortative with respect to the eigenvector and closeness centrality metrics for the actual US States network graph.

Thus, based on the results obtained for the node-level metrics, we could conclude that the degree sequence of the US States network graph would be sufficient to generate random network instances that exhibit strong-very strong positive levels of correlation with respect to all the four centrality metrics. On the other hand, with respect to the other node-level metrics (like Eccentricity, Maximal Clique Size and Local Clustering Coefficient) as well as for all the network-level metrics (other than Degree centrality and Spectral radius ratio for node degree), we could conclude that the degree sequence of the US States network graph would alone not be sufficient to generate random network instances that exhibit comparable values for these metrics.

## 6. Related Work

Very few works have been conducted on network graphs related to the US. We review these works below: Fogarty et al. (2008) conducted a network analysis-based study on the hurricanes that made landfalls in the US from 1851 to 2008. A set of 23 non-overlapping regions (nodes) of the US that were affected with at least one hurricane were identified; two nodes were linked with an edge if at least one hurricane impacted the regions corresponding to both of them. One of the interesting conclusions from this study was that regions (like Louisiana) with a high occurrence rate of hurricanes had a low connectivity with the rest of the regions; on the other hand, regions with high connectivity (like Virginia) had a low occurrence rate. Several similarities have been observed between the hurricane landfall network by Fogarty et al (2008) and the US states network graph studied in this paper. For both the networks, the betweenness centrality metric exhibited a power-law distribution and the closeness centrality metric exhibited a uniform distribution with narrow range of values. While the average local clustering coefficient of the nodes in the landfall network was 0.46, the average local clustering coefficient of the nodes in the US states network graph is slightly larger (0.52). The diameter values for the network graphs are proportional: we observe a diameter of 10 for the US states network graph of 49 nodes and a diameter of 5 for the landfall network of 23 nodes. However, the two networks differ with respect to the degree centrality metric: we observe a clear bi-modal degree distribution for the US states network graph and no such distinct distribution could be attributed for the degree centrality metric in the landfall network. Though the hurricane landfall network and the US States network shared several similarities (as mentioned above), it must be remembered that the hurricane landfall network was constructed by cumulatively considering the landfall of hurricanes over a longer period of time (for about 150 years). We anticipate the results for the node-level and network-level metrics to appreciably differ for the two networks if the landfall network is constructed for a particular year or over a shorter time period.

Lin et al. (2014) conducted a network analysis of food flows within the US and had the following results: The distributions for the degree centrality and betweenness centrality were observed to be normal and Weibull (Balakrishnan, & Nevzorov, 2003) in nature. A power-law relationship (Balakrishnan, & Nevzorov, 2003) existed between the degree centrality and betweenness centrality metrics, indicating a vulnerability to the disturbance of key nodes. On the other hand, we did not observe a power-law relationship between degree and betweenness centrality for the US States network graph; even vertices with moderate-high degree had a low betweenness centrality. Lyte et al. (2015) conducted a citation network-based analysis of the different sections that fall under the 52 titles of United States Code; each section is a node and there exists a directed edge from one section to another section if the former cites the latter. The betweenness and eigenvector centrality metrics were used in this study to identify major pathways of references from one section to another. The modularity-based Louvain community detection algorithm (Blondel et al., 2008) was used to identify communities of sections that had similarities with respect to concepts and codes. It was observed that though sections under two or more related titles formed a single community, most of the communities detected were a collection of sections under a particular title. For the US States network graph, the communities detected using the Louvain algorithm were similar to the regional divisions used by the United States Census Bureau.

Cheung and Gunes (2012) conducted a complex network analysis study of the US air transportation network as of 2011 and compared it with the networks that existed in 1991 and 2001. Their study revealed no major changes in the features (like centrality and connectivity of the airports) of the air transportation networks that evolved with time (with increase in the number of airports and flight connections). A critical finding from the study was that the US air transportation network of 2011 has been identified to be more vulnerable to airport closures than it was in the past. The degree distribution of the 2011 US air transportation network only follows a partial Power-law (i.e., the distribution exhibited Power-law only after a degree value > 1), unlike the world-wide air transportation network that follows Power-law starting from degree value of 1 (Guimera, 2005). Random network instances (generated using the configuration model) of the US States network graph exhibited strong positive correlation with respect to the centrality metrics, but were observed to be relatively more bipartite, modular and robust to disconnection.

## 7. Summary and Conclusions

Our high-level contribution in this paper is to illustrate complex network analysis of a connected graph of the states within a country at node-level and network-level as well as propose a normalization-based approach to comprehensively rank the vertices (more likely to be tie-free) in a network graph based on the centrality metrics. We implemented the algorithms to compute a suite of node-level and network-level metrics and ran them on the US States network graph. We summarize the results and key observations as follows: (i) The state of Missouri is the top-ranked node with respect to all the commonly studied centrality metrics such as degree, betweeenness, closeness and eigenvector centralities. This is vindicated with several airlines (like American Airlines, Southwest Airlines, etc) choosing the city of Missouri as one of their primary hubs over the past two decades. (ii) The degree distribution appears to mimic a bi-modal Poisson distribution, while the betweenness centrality (BWC) exhibits a Power-law style distribution. (iii) There exists a maximum clique of size 4 involving the states of Arizona, Colorado, New Mexico and Utah; the rest of the states (except Maine) are part of maximal cliques of size 3. (iv) The state of Idaho has the lowest non-zero local clustering coefficient, indicating that the state is the most critical state with respect to facilitating communication between its neighboring states. (v) The radius, diameter and average path length are 5, 10 and 3.94 respectively. The states of Ohio and West Virginia form the "center" of the graph with an eccentricity corresponding to the radius of the graph (these states are at most 5 hops away from any other state in the graph). The states of Arizona, California, Maine, Montana and North Dakota have an eccentricity corresponding to the diameter of the graph (these states could be as large as 10 hops away to one or more states in the graph). More than 65% of the vertices have an eccentricity of 8 or above. (vi) The bipartivity index of the graph is 0.66 with 32% frustrated edges. (vii) The algebraic connectivity of the network graph is 0.0973 (indicating low robustness) and the spectral radius ratio for node degree is 1.24 (moderately high for a Poisson network, vindicating the bi-modal degree distribution of the vertices). (viii) The modularity score of the graph is 0.58 with a total of six non-overlapping communities of states, closely resembling the regional classification of the states. (ix) The network has been observed to be relatively more assortative with respect to eigenvector and closeness centralities; whereas the degree-based and BWC-based approximations to the minimum connected dominating sets are of the smallest size. (x) The Configuration model-based study of the US States network graph indicated that the degree sequence alone was sufficient to generate random network instances that exhibited strong-very strong levels of positive correlation for the centrality metrics, but the degree sequence was not sufficient to observe such a strong correlation for the other node-level metrics and comparable values for the network-level metrics. The random network instances of the US States network graph were observed to be relatively more robust to network disconnection, more bipartite and more modular. Thus, even though it might look like some states may have a common border by chance (especially, if the common border is over a smaller area), the above results (especially those from assortativity analysis and the configuration model-based study) indicate that the network of US states is very much different from a random network.

We have also proposed a normalization-based approach to arrive at a (possibly tie-free) ranking of the vertices based on their comprehensive centrality scores determined as a weighted average of the normalized scores of the individual centrality metrics. We also show how to identify the centrality metric whose normalized individualized scores and ranking of the vertices is relatively the closest to the normalized comprehensive centrality (NCC) scores and the ranking of the vertices based on the NCC scores. Considering the results plotted in Figures 12-(a) through 12-(d) and Figures 13-(a) through 13-(d), it appears that the Eigenvector Centrality metric (that consistently incurs the second smallest RMSD values with respect to both the normalized centrality scores and the numerical ranking of the vertices) could be relatively the best metric that could be used to obtain a comprehensive centrality-based ranking of the vertices in the US States network graph. A similar approach could be used to identify a centrality metric that could be considered the candidate metric to claim a comprehensive centrality-based ranking of the vertices in other real-world network graphs and synthetic graphs generated from theoretical models.

To the best of our knowledge, we have not come across a paper that comprehensively analyzes a suite of node-level and network-level metrics for any real-world network and one especially based on the states within a country. The approach taken and the metrics evaluated in this paper could have several applications: For example, we could identify the states that are most the central states as well as identify the states that could form a connected backbone and geographically well-connected to the rest of the states within a country and use this information to design the road/rail transportation networks; we could identify the states that could be clustered to a particular geographical region within a country and use this information for region-based analysis and etc. For countries with a reasonably larger area and an appreciable number of states, each state (except those in the corners of the country) typically shares border with a similar number of states. Hence, we anticipate the distribution of values for the node-level metrics to be about the same for several other countries too. We thus opine the paper to serve as a model for anyone interested in analyzing a connected graph of the states within a country from a Network Science perspective.

## Acknowledgments

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

## References

Balakrishnan, N., & Nevzorov, V. B. (2003). *A Premier on Statistical Distributions*. (1st ed.) Wiley-Interscience.

Balanda, K. P., & MacGillivray, H. L. (1998). Kurtosis: A Critical Review. *The American Statistician*, *42*(2), 111-119. http://dx.doi.org/10.2307/2684482.

Barabasi, A. L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, *286*(5439), 509-512. http://dx.doi.org/10.1126/science.286.5439.509.

Benson, S. J. (2008). *Explorer's Guide The Four Corners Region: Where Colorado, Utah, Arizona & New Mexico Meet: A Great Destination*. (1$^{st}$ ed.) Countryman Press.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics*: *Theory and Experiment*, P10008, 1-11. http://dx.doi.org/10.1088/1742-5468/2008/10/P10008.

Brandes, U. (2001). A Faster Algorithm for Betweenness Centrality. *The Journal of Mathematical Sociology*, *25*(2), 163-177. http://dx.doi.org/10.1080/0022250X.2001.9990249.

Cherven, K. (2015). *Mastering Gephi Network Visualization*. (1$^{st}$ ed.) Packt Publishing.

Cheung, D. P., & Gunes, M. H. (2012). *A Complex Network Analysis of the United States Air Transportation*. Paper presented at the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Istanbul, Turkey. http://dx.doi.org/10.1109/ASONAM.2012.116.

Chung, L. L. F. (2006). *Complex Graphs and Networks*. (1$^{st}$ ed.) American Mathematical Society.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms*. (3rd ed.) MIT Press.

Daniel, W. W. (2000). *Applied Nonparametric Statistics*. (2nd ed.) Cengage Learning.

Ding, Y. (2011). Scientific Collaboration and Endorsement: Network Analysis of Coauthorship and Citation Networks. *Journal of Informetrics*, *5*(1), 187-203. http://dx.doi.org/10.1016/j.joi.2010.10.008.

Erdos, P., & Renyi, A. (1959). On Random Graphs I. *Publicationes Mathematicae*, *6*, 290-297.

Estrada, E., & Rodriguez-Velazquez, J. A. (2005). Spectral Measures of Bipartivity in Complex Networks. *Physical Review E*, *72*(4), 046105. https://doi.org/10.1103/PhysRevE.72.046105.

Evans, J. D. (1995). *Straightforward Statistics for the Behavioral Sciences*. (1st ed.) Brooks Cole Publishing Company.

Fiedler, M. (1973). Algebraic Connectivity of Graphs. *Czechoslovak Mathematical Journal*, *23*(2), 298-305.

Fogarty, E. A., Elsner, J. B., Jagger, T. H., & Tsonis, A. A. (2008). Network Analysis of U. S. Hurricanes. *Springer Hurricanes and Climate Change*, 153-167. http://dx.doi.org/10.1007/978-0-387-09410-6_9.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph Drawing by Force-Directed Placement. *Software - Practice & Experience*, *21*(11), 1129-1164.

Ghali, N., Panda, M., Hassanien, A. E., Abraham, A., & Snasel, V. (2012). Social Network Analysis: Tools, Measures and Visualization. *Computational Social Networks*, 3-23. http://dx.doi.org/10.1007/978-1-4471-4054-2_1.

Girvan, M., & 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*, *19*(12), 7821-7826. http://dx.doi.org/10.1073/pnas.122653799.

Guha. S., & Khuller, S. (1998). Approximation Algorithms for Connected Dominating Sets. *Algorithmica*, *20*(4), 374-387. http://dx.doi.org/10.1007/PL00009201.

Guimera, R., Mossa, S., Turtschi, A., & Amaral, L. A. N. (2005). The World-Wide Air Transportation Network: Anomalous Centrality, Community Structure, and Cities' Global Roles. *Proceedings of the National Academy of Sciences*, *102*(22), 7794-7799. http://dx.doi.org/10.1073/pnas.0407994102.

Lay, D. C., Lay, S. R., & McDonald, J. J. (2015). *Linear Algebra and its Applications*. (5th Ed.) Pearson.

Lin, X., Dang, Q., & Konar, M. (2014). A Network Analysis of Food Flows within the United States of America. *Environmental Science & Technology*, *48*(10), 5439-5447. http://dx.doi.org/10.1021/es500471d.

Lyte, A., Slater, D., & Michel, S. (2015). Network Measures of the United States Code. Technical Report, MITRE.

Ma, X., & Gao, L. (2012). Biological Network Analysis: Insights into Structure and Functions. *Briefings in Functional Genomics*, *11*(6), 434-442, November 2012. https://doi.org/10.1093/bfgp/els045.

Meghanathan, N. (2014a). 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. http://dx.doi.org/10.1109/UNESST.2014.8.

Meghanathan, N. (2014b). Centrality-based Connected Dominating Sets for Complex Network Graphs. *International Journal of Interdisciplinary Telecommunications and Networking*, *6*(2), 1-19. http://dx.doi.org/10.4018/ijitn.2014040101.

Meghanathan, N. (2015a). Exploiting the Discriminating Power of the Eigenvector Centrality Measure to Detect Graph Isomorphism. *International Journal in Foundations of Computer Science and Technology*, *5*(6), 1-13. http://dx.doi.org/10.5121/ijfcst.2015.5601.

Meghanathan, N. (2015b). Distribution of Maximal Clique Size of the Vertices for Theoretical Small-World Networks and Real-World Networks. *International Journal of Computer Networks and Communications*, *7*(4), 21-41. http://dx.doi.org/10.5121/ijcnc.2015.7402.

Meghanathan, N. (2016a). Maximal Assortative Matching for Complex Network Graphs. *Journal of King Saud University: Computer and Information Sciences*, *28*(2), 230-246. http://dx.doi.org/10.1016/j.jksuci.2015.10.004.

Meghanathan, N. (2016b). On the Conduciveness of Random Network Graphs for Maximal Assortative or

Maximal Dissortative Matching. *Computer and Information Science*, *9*(1), 21-30. http://dx.doi.org/10.5539/cis.v9n1p21.

Meghanathan, N. (2016c). On the Sufficiency of using the Degree Sequence of the Vertices to Generate Random Networks Corresponding to Real-World Networks. *Polibits: Research Journal on Computer Science and Computer Engineering with Applications*, *53*(1), 5-21. http://dx.doi.org/10.17562/PB-53-1.

Newman, M. (2006). Modularity and Community Structure in Networks. *Proceedings of the National Academy of Sciences*, *103*(23), 8577-8582. http://dx.doi.org/10.1073/pnas.0601602103.

Newman, M. (2010). *Networks*: *An Introduction*. (1st ed.) Oxford University Press.

Pattabiraman, B., Patwary, M. A., Gebremedhin, A. H., Liao, W-K., & Choudhary, A. (2013). *Fast Problems for the Maximum Clique Problem on Massive sparse Graphs*. Paper presented at the 10th International Workshop on Algorithms and Models for the Web Graph, Cambridge, MA, USA. http://dx.doi.org/10.1007/978-3-319-03536-9_13.

Triola, M. F. (2012). *Elementary Statistics*. (12th ed.) Pearson.

Zhao, D., & Strotmann, A. (2015). *Analysis and Visualization of Citation Networks*. (1st ed.) Morgan & Claypool Publishers.

## Copyrights