

Md Atiqur Rahman and Natarajan Meghanathan

## 33.1 Introduction

Complex network analysis is fundamental to the understanding and modeling of relationships in several domains such as social networks [1], communication networks [2], biological networks [3], citation networks [4], cyber-attack networks [5] etc. Centrality metrics are the most common node-level metrics and quantify the topological importance of a node (vertex) with respect to one or more aspects [6]. There are two major categories of centrality metrics: neighborhood-based and shortest path-based. While the degree (DEG) and eigenvector centrality (EVC) metrics are considered representative metrics for the neighborhood-based category, the betweenness (BWC) and closeness (CLC) centrality metrics are considered representative metrics for the shortest path-based category. The DEG of a vertex is a measure of the number of neighbors of the vertex. The EVC of a vertex [7] is a measure of the degree of the vertex as well as the degrees of its neighbors. The BWC of a vertex [8] is a measure of the fraction of the shortest paths between any two vertices that go through the vertex. The CLC of a vertex [9] is a measure of the sum of the shortest path distances from the vertex to every other vertex. For more information about these centrality metrics, the interested reader is referred to [6–9]. Throughout the paper, the terms ‘node’ and ‘vertex’, ‘link’ and ‘edge’, ‘network’ and ‘graph’, ‘cluster’ and ‘community’ are used interchangeably. They mean the same.

Community detection is a classical problem in complex network analysis wherein one or more clusters of vertices are identified based on the topological distribution of the vertices and the edges connecting the vertices [6]. The primary focus of the community detection algorithms (e.g., [10]) has been to determine highly modular communities such that the intra-cluster density is as high as possible and the inter-cluster density is as low as possible [11]. With the above approach for clustering in complex networks, two vertices are likely to belong to the same cluster only if they have an edge between them or have a multi-hop shortest path of fewer intermediate edges. Even within a cluster of high density, it is possible that vertices significantly differ with respect to the values for one or more node-level metrics (see Sect. 33.2 for a motivating example) if the extent of similarity of the vertices (with respect to one or more node-level metrics) is not the primary criteria for cluster formation.

Our research objective is to quantify the extent to which vertices in a complex network are logically clusterable on the basis of the similarity among the values incurred for one or more node-level metrics. The nodes in such a logical cluster need not be connected, but would have very similar values for the node-level metrics. Identification of such logical clusters would be very informative for several network domains. For example: in an organizational network, a logical cluster of employees with similar profiles (but in different departments) would help in forming an integration team that could coordinate across departments without any ego differences.

---

Md. A. Rahman  
Computational and Data-Enabled Science and Engineering, Jackson  
State University, Jackson, MS, USA  
e-mail: [md.a.rahman@students.jsums.edu](mailto:md.a.rahman@students.jsums.edu)

N. Meghanathan (✉)  
Department of Computer Science & Engineering, Jackson State  
University, Jackson, MS, USA  
e-mail: [natarajan.meghanathan@jsums.edu](mailto:natarajan.meghanathan@jsums.edu)

In a health information network, a quantitative measure of clusterability of patients with similar health parameters is critical to decide whether treatment plans need to be individualized or can be generalized at the cluster-level. The Hopkins Statistic measure (ranges from 0 to 1) has been traditionally used to assess the clusterability of a dataset [12]. We use the Hopkins Statistic to assess the similarity-based clusterability of the vertices in a complex network.

The rest of the paper is organized as follows: Section 33.2 motivates the proposed research for similarity-based logical clustering with an illustrative example. Section 33.3 illustrates the computation of the Hopkins Statistic on a coordinate system of the normalized centrality values for the example graph of Sect. 33.2. Section 33.4 presents the Hopkins Statistic values computed for 47 real-world networks (of diverse degree distributions) and evaluates the logical clusterability of the vertices on the basis of the similarity with respect to the neighborhood-based centrality metrics vs. the shortest path-based centrality metrics. Section 33.5 reviews related work on similarity assessment and clustering in complex networks, and highlights our contribution. Section 33.6 concludes the paper and outlines plans for future work.

### 33.2 Proposed Approach and Example

Our proposed approach to quantitatively assess similarity-based clusterability is as follows: Let  $K$  be the number of centrality metrics considered for the analysis. We determine the centrality values of the vertices in the complex network with respect to each of these metrics. We then individually normalize the centrality values of the vertices for each metric. The normalization of the centrality values with respect to a metric is done using the square root of the sum of the squares of the centrality values. We distribute the vertices in a  $K$ -dimensional coordinate system (wherein the range for each dimension is from 0 to 1) such that the coordinate for a vertex is a  $K$ -tuple, with an entry for each centrality metric. We refer to the above distribution as the logical topology of the vertices. If two vertices have similar values for the centrality metrics, then they should be closer to each other in the logical topology. There could exist one or more logical clusters of such similar vertices so that the Euclidean distance between vertices within a cluster is much smaller than the Euclidean distance between vertices in two different clusters. If the centrality values of the vertices are not similar to each other, then the vertices could be further away from each other in the logical topology and such a topology would not be effectively clusterable. The Hopkins Statistic is an effective measure to quantify the extent to which vertices could be clustered based on the similarity of their centrality values. The larger the Hopkins Statistic, the more effective is the clusterability of the vertices on the basis of similarity in their centrality values.

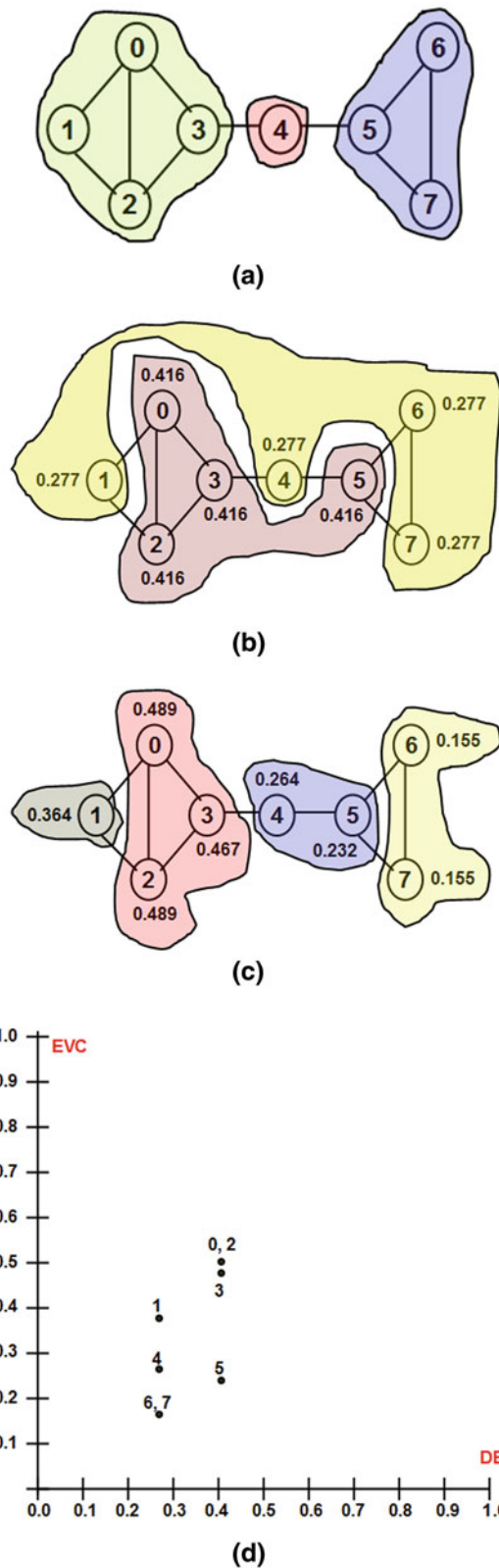
Figure 33.1a presents the communities determined by the classical Girvan-Newman algorithm [10] on a toy example graph. We observe the nodes in the individual clusters of Fig. 33.1a to be more connected to each other than to the nodes in the other clusters (i.e., a larger intra-cluster density and a lower inter-cluster density). However, from a centrality point of view, we observe the nodes in the individual clusters to be much different from each other (see Fig. 33.1b and c). Figure 33.1b and c respectively present a logical clustering of the vertices based on the similarity of the normalized values of the DEG and EVC metrics. The vertices within such logical clusters need not be physically connected to each other (Fig. 33.1b) or even if connected, the logical clusters need not have a high modularity (Fig. 33.1c), but would have similar (either identical or very close) values for the centrality metric considered.

If two or more centrality metrics are to be considered together for similarity assessment, then we distribute the vertices in a coordinate system of the normalized centrality values (ranging from 0 to 1 for each dimension) and assess the clusterability of the vertices based on their proximity to each other. Figure 33.1d presents one such distribution of the vertices based on a coordinate system of the normalized DEG and EVC values. Though the exact clusters depend on the clustering algorithm used, we observe vertex 5 to be far away from the rest of the vertices in the normalized (DEG, EVC) co-ordinate system and such a conclusion cannot be arrived at in Fig. 33.1a, b or c. Likewise, vertex 1 gets grouped with vertices 0, 2 and 3 under the traditional approach (see Fig. 33.1a), but is reasonably far away from them in the (DEG, EVC) co-ordinate system.

### 33.3 Hopkins Statistic Measure

We use the Hopkins Statistic measure [12] as the basis to quantify the similarity-based clusterability of the vertices in a complex network. The Hopkins Statistic (ranges from 0 to 1) is traditionally used to assess the clustering tendency of any dataset. The idea behind the Hopkins Statistic is as follows: Given a dataset of size  $n$ , we pick  $m$  samples (constituting the set  $X$  such that  $m \ll n$ ) from this dataset as well as generate  $m$  uniformly random samples (constituting the set  $Y$ ) exhibiting the same variation as that of the given dataset. We determine the sums of the distances of the samples in the sets  $X$  and  $Y$  to the nearest data points (i.e., one nearest data point for each sample) in the given dataset. The Hopkins Statistic is the ratio of the sum of the nearest neighbor distances of the samples in the set  $Y$  to the sum of the nearest neighbor distances of the samples in the sets  $X$  and  $Y$ . If the data points in the given dataset are randomly distributed, then the sums of the nearest neighbor distances for the samples in the sets  $X$  and  $Y$  would be approximately the same, and the Hopkins Statistic is expected to be 0.5. If the data points in the given

**Fig. 33.1** Example to illustrate the clustering of the vertices based on the traditional approach vs. the normalized centrality-based coordinate system. **(a)** Physical clusters identified by the traditional community detection approach. **(b)** Logical clustering of the vertices based on the similarity with respect to degree centrality. **(c)** Logical clustering of the vertices based on the similarity with respect to eigenvector centrality. **(d)** Distribution of vertices in the logical topology respect to degree and eigenvector centrality metrics



dataset are highly clusterable, then the sum of the nearest neighbor distances for the samples in the set X would be very negligible compared to the sum of the nearest neighbor

distances for the samples in the set Y; hence, the Hopkins Statistic is expected to be close to 1.0 for a highly clusterable dataset. If the data points in the original dataset are uniformly

distributed, then the sum of the nearest neighbor distances for the samples in the set X would be significantly larger than the sum of the nearest neighbor distances for the samples in the set Y; the Hopkins Statistic in such a case would be closer to 0. A value of 0.75 or higher for the Hopkins Statistic indicates a clustering tendency at the 90% confidence level [5] [13–15].

We now illustrate the procedure to compute the Hopkins Statistic measure for the logical topology of the vertices illustrated in Fig. 33.1d. The distance between two vertices in the logical topology is computed as the Euclidean distance between their coordinates (represented by the normalized centrality values with respect to degree and eigenvector centrality). There are a total of eight vertices ( $n = 8$ ) in the toy example graph of Fig. 33.1. Let the parameter ‘ $m$ ’ be 5. That is, we randomly pick five of the eight vertices from the graph and constitute the set X. Let the five vertices picked be {0, 3, 4, 6, 7}, and the set X would comprise of their corresponding (DEG, EVC) values as coordinates: {(0.416, 0.489), (0.416, 0.467), (0.277, 0.264), (0.277, 0.155), (0.277, 0.155)}. From Fig. 33.1b and c, we could identify the range of the DEG and EVC values to be [0.277, . . . , 0.416] and [0.155, . . . , 0.489] respectively. While the normalized DEG values for the vertices in the original dataset are either 0.277 or 0.416, the normalized EVC values do not appear to follow any particular pattern. Hence, to generate the set Y with five data points, the DEG values need to be either 0.277 or 0.416 (we randomly choose one of these two values for each data point) and the EVC values need to be uniform randomly distributed in the range [0.155, . . . , 0.489]. Accordingly, the five samples constituting the set Y are generated to be: {(0.416, 0.168), (0.277, 0.328), (0.416, 0.479), (0.416, 0.183), (0.277, 0.217)}.

Figure 33.2 summarizes all the calculations. The sums of the nearest distances to a vertex for the samples in the sets X and Y are respectively 0.122 and 0.276; the Hopkins Statistic for the (DEG, EVC) coordinate system is  $0.276 / (0.276 + 0.122) = 0.69$ . Since it is relatively closer to

0.5, we could conclude that the vertices in the given complex network graph are not very much clusterable on the basis of the similarity with respect to the DEG and EVC values.

### 33.4 Similarity-Based Clusterability of Real-World Networks

We applied the proposed approach to a suite of 47 real-world networks of diverse degree distributions and computed the values for the Hopkins Statistic for the logical topologies of the vertices with respect to the neighborhood-based DEG and EVC metrics and the shortest path-based BWC and CLC metrics. The real-world networks (for more details, refer to [16]) considered fall under one of these domains (the number of networks under each category is indicated within the parentheses): I. Acquaintance network (12), II. Friendship network (9), III. Co-appearance network (6), Employment network (4), Citation network (3), Collaboration network (3), Literature network (3), Biological network (3), Political network (2), Game network (2), Transportation network (1), Geographical network (1) and Trade network (1). Table 33.1 presents the Hopkins Statistic values for each network with respect to the (DEG, EVC) and (BWC, CLC) measures along with the number of nodes and edges as well as the average degree ( $k_{avg}$ ) and spectral radius ratio for node degree [17] for each of these networks. The spectral radius ratio for node degree ( $\lambda_{sp}$ ) is a measure of the variation in node degree.

Figure 33.3 plots the Hopkins Statistic values for the two coordinate systems/logical topologies: if a data point is below the diagonal line, then the Hopkins Statistic value for the (DEG, EVC) coordinate system is relatively larger; if a data point is above the diagonal line, then the Hopkins Statistic value for the (BWC, CLC) coordinate system is relatively larger. Table 33.1 highlights the Hopkins Statistic values that are (overall) relatively larger for a real-world network with respect to the two coordinate systems. From both Fig. 33.3 and Table 33.1, we could come to the following conclusions:

**Fig. 33.2** Calculations for the Hopkins Statistic for (DEG, EVC)-Similarity based Clusterability for the graph of Fig. 33.1

Vertex ID	DEG	EVC	Vertex in Set X	DEG	EVC	Nearest Distance	Nearest Vertex
0	0.416	0.489	0	0.416	0.489	0.000	2
1	0.277	0.364	3	0.416	0.467	0.022	2
2	0.416	0.489	4	0.277	0.264	0.100	1
3	0.416	0.467	6	0.277	0.155	0.000	7
4	0.277	0.264	7	0.277	0.155	0.000	6
5	0.416	0.232	Sum = 0.122				
6	0.277	0.155					
7	0.277	0.155					

Data Point in Set Y	DEG	EVC	Nearest Distance	Nearest Vertex
1	0.416	0.168	0.064	5
2	0.277	0.328	0.036	1
3	0.416	0.479	0.080	0
4	0.416	0.183	0.049	5
5	0.277	0.217	0.047	4
Sum = 0.276				

<b>Hopkins Statistic</b>
0.276
= -----
(0.276 + 0.122)
= 0.693

**Table 33.1** Hopkins statistic values with respect to the neighborhood-based (DEG, EVC) metrics vs. shortest path-based (BWC, CLC) metrics

#	Net.	Network Name	# nodes	# edges	$k_{avg}$	$\lambda_0$	(DEG, EVC)	(BWC, CLC)
1	ADJ	Word Adjacency Net <sup>III</sup>	112	425	7.6	1.7	0.92	0.84
2	AKN	Anna Karenina Net <sup>III</sup>	140	494	7.1	2.5	0.93	0.92
3	JBN	Jazz Band Net <sup>IV</sup>	198	2742	27.7	1.4	0.92	0.96
4	CEN	C. Elegans Neural Net <sup>VIII</sup>	297	2148	14.4	1.7	0.96	0.98
5	CLN	Centrality Literature Net <sup>V</sup>	118	613	10.4	2.0	0.93	0.85
6	CGD	Citation Graph Dr. Net <sup>V</sup>	259	640	4.1	2.2	0.84	0.95
7	CFN	Copperfield Net <sup>III</sup>	89	407	9.1	1.8	0.96	1.00
8	DON	Dolphin Net <sup>I</sup>	62	159	5.1	1.4	0.79	0.73
9	DRN	Drug Net <sup>I</sup>	212	284	1.9	2.8	0.89	0.96
10	DLN	Dutch Lit. 1976 Net <sup>VIII</sup>	37	81	4.4	1.5	0.78	0.64
11	ERD	Erdos Collab. Net <sup>VI</sup>	433	1314	6.1	3.0	0.94	0.99
12	FMH	Faux Mesa High Sc. Net <sup>II</sup>	147	202	1.9	2.8	0.86	0.96
13	FHT	Friendship Hi-Tec Firm <sup>II</sup>	33	91	5.1	1.6	0.81	0.77
14	FTC	Flying Teams Cad. Net <sup>IV</sup>	48	170	7.1	1.2	0.77	0.76
15	FON	US Football Net <sup>X</sup>	115	613	10.7	1.01	0.94	0.71
16	CDF	College Dorn Fraternity <sup>I</sup>	58	967	33.3	1.1	0.91	0.79
17	GD96	Graph Drawing '96 Net <sup>V</sup>	180	228	2.5	2.4	0.96	0.95
18	MUN	Marvel Universe Net <sup>III</sup>	167	301	3.6	2.5	0.93	0.98
19	GLN	Graph Glossary Net <sup>III</sup>	67	118	3.3	2.0	0.87	0.93
20	HTN	Hypertext 2009 Net <sup>I</sup>	115	2164	38.3	1.2	0.93	0.84
21	HCN	Huckleberry Co. Net <sup>III</sup>	76	302	7.9	1.7	0.93	0.99
22	ISP	Infec. Socio-Patt. Net <sup>I</sup>	309	1924	12.4	1.7	0.91	0.92
23	KCN	Karate Club Net <sup>I</sup>	34	78	4.6	1.5	0.87	0.80
24	KFP	Korea Family Plan. Net <sup>I</sup>	37	85	4.3	1.7	0.85	0.73
25	LMN	Les Miserables Net <sup>III</sup>	77	254	6.6	1.8	0.87	0.96
26	MDN	Macaque Dom. Net <sup>VIII</sup>	62	1167	37.6	1.04	0.90	0.75
27	MTB	Madrid Train Bomb. Net <sup>I</sup>	64	295	9.2	1.9	0.89	0.86
28	MCE	Manufact. Comp. Empl <sup>IV</sup>	77	1549	40.2	1.1	0.89	0.84
29	MSJ	Social Net. Journal Net <sup>VI</sup>	475	625	2.6	3.5	0.99	0.93
30	AFB	Author Facebook Net <sup>II</sup>	171	940	10.0	2.3	0.94	0.96
31	MPN	Mexican Pol. Elite Net <sup>IX</sup>	35	117	6.7	1.2	0.74	0.65
32	MMN	ModMath Net <sup>II</sup>	30	61	4.1	1.6	0.74	0.62
33	PBN	US Politics Books Net <sup>VIII</sup>	105	441	8.4	1.4	0.84	0.85
34	PSN	Primary Sc. Contact Net <sup>I</sup>	238	5539	46.5	1.2	0.93	0.85
35	PFN	Prison Friendship Net <sup>II</sup>	67	142	4.2	1.3	0.82	0.80
36	SJN	San Juan Sur Family Net <sup>I</sup>	75	155	4.1	1.3	0.78	0.80
37	SDI	Scotland Cor. Interlock <sup>IV</sup>	230	359	2.9	1.9	0.96	0.98
38	SPR	Senator Press Rel. Net <sup>IX</sup>	92	477	10.4	1.6	0.91	0.78
39	SWC	Soccer World Cup Net <sup>X</sup>	35	118	6.7	1.4	0.85	0.70
40	SSM	Sawmill Strike Com. Net <sup>I</sup>	24	38	3.2	1.2	0.59	0.75
41	TEN	Taro Exchange Net <sup>I</sup>	22	39	3.5	1.06	0.67	0.56
42	TWF	Teenage Fem. Fr. Net <sup>II</sup>	47	77	3.3	1.5	0.82	0.70
43	UKF	UK Faculty Friend Net <sup>II</sup>	83	578	14.2	1.3	0.86	0.88
44	APN	US Airports 1997 Net <sup>III</sup>	332	2126	14.3	3.2	0.96	0.94
45	RHF	Residence Hall Fr. Net <sup>II</sup>	217	1839	16.9	1.3	0.92	0.91
46	WSB	Windsurfers Beach Net <sup>II</sup>	43	336	15.6	1.2	0.84	0.78
47	WTN	World Trade Met. Net <sup>VIII</sup>	80	875	21.8	1.5	0.93	0.88

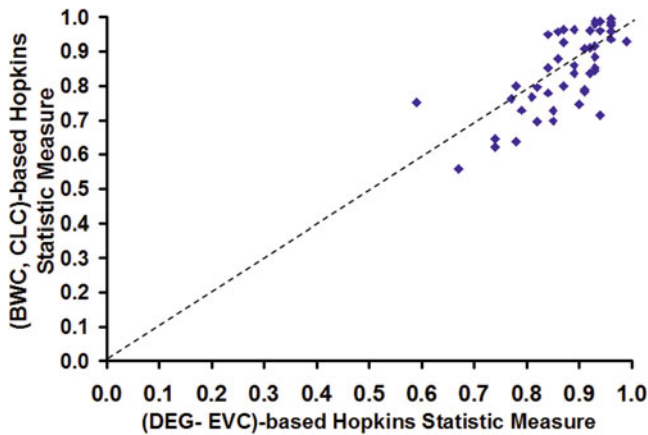
The median of the Hopkins Statistic values for the (DEG, EVC) and (BWC, CLC)-based logical topologies are 0.89 and 0.85 respectively. Hence, majority of the real-world networks are effectively clusterable with respect to both the neighborhood and shortest path-based centrality metrics. On a relative basis: out of the 47 real-world networks, for 29 networks (i.e., for about 60% of the real-world networks): the (DEG, EVC)-based Hopkins Statistic values are larger. Also, the Hopkins Statistic values for the (DEG, EVC)-based logical topologies are relatively closer to 1.0 compared to the Hopkins Statistic values for the (BWC, CLC)-based logical topologies. We could thus conclude that real-world networks are more effectively clusterable with respect to the neighborhood-based centrality metrics compared to shortest path-based metrics.

Figure 33.4 plots the distribution of the Spectral radius ratio for node degree vs. the Hopkins Statistic measure values

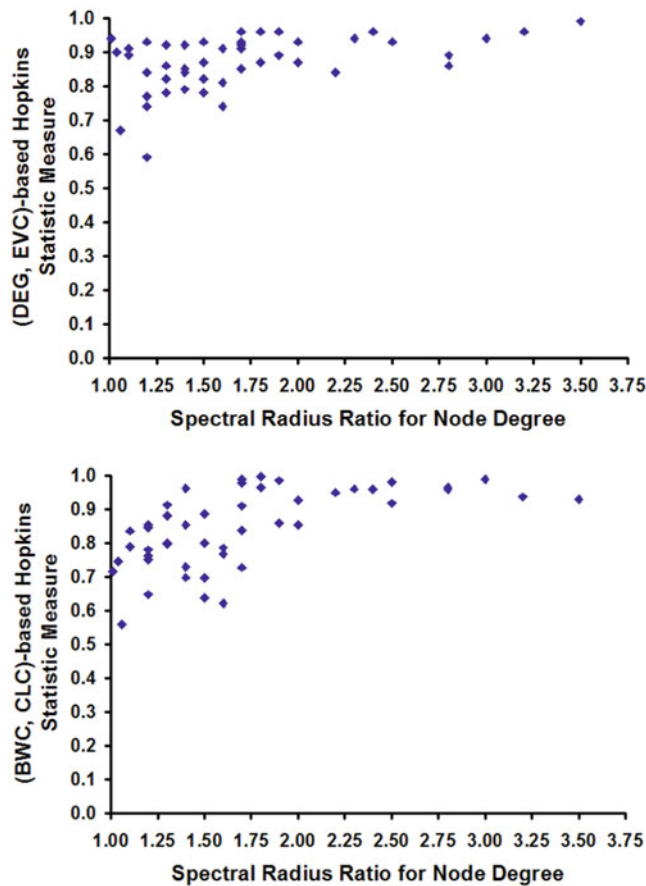
for the (DEG, EVC) and (BWC, CLC)-based logical topologies. We observe for both the coordinate systems, the logical topologies are more effectively clusterable with increase in the spectral radius ratio for node degree. This implies: scale-free networks (larger values for the spectral radius ratio for node degree) are more effectively clusterable with respect to both categories of centrality metrics. Whereas, the random networks (that have lower values for the spectral radius ratio for node degree) are (relatively) less effectively clusterable with respect to both categories of centrality metrics, especially the shortest path-based centrality metrics.

### 33.5 Related Work and Our Contribution

Clustering in complex networks has been traditionally on the notion of distance between the nodes [13]. Nodes that are



**Fig. 33.3** Comparison of the Hopkins Statistic values for the (DEG, EVC)-based logical topologies vs. (BWC, CLC)-based logical topologies of the real-world networks



**Fig. 33.4** Spectral radius ratio for node degree vs. the Hopkins statistic measure values

closer to each other are preferred for being grouped within a cluster. Various forms of distance measures (Euclidean, Manhattan, Minkowski, Chebyshev, Mahalanobis, etc) have been used in the literature [14]. Similarity-based clustering has been typically considered only for non-network data [15, 18], broadly as a problem in data mining. Given a distribution

of data points, data points that are similar with respect to one or more parameters tend to be clustered together [19, 20]. The typical approach [19] is to associate a multi-dimensional vector to each data point, measure the pair-wise similarity (using functions such as cosine similarity [6], Jaccard Index [19], SimRank [21], PathSim [22]) or the distances between the data points (using some distance function), and then cluster the data points that are similar and closer to each other. Our research proposes a combination of the above two approaches. We first determine the values for the centrality metrics (node-level parameters) of the nodes in a complex network, then treat the nodes as data points with an associated multi-dimensional vector of the normalized values of the centrality metrics, and finally analyze the clusterability of these data points.

The proposed Hopkins Statistic-based clusterability assessment of node similarity could be the first step in pursuit of developing a network-level similarity index for complex networks. If the Hopkins Statistic measure is high, there could exist one or more clusters of similar nodes (with respect to the centrality metrics) in the network. Though there could exist two or more such clusters in the network (with the nodes in one cluster significantly different from the nodes in the other clusters with respect to the centrality metrics), a lower value for the Hopkins Statistic measure is definitely an indication that the nodes in the logical topology of centrality-based coordinates are more randomly distributed and not very similar to each other (i.e., not clusterable). In other words, the computation of the Hopkins Statistic measure could be used as a prerequisite step for any algorithm to quantify/assess the network-wide similarity of all the nodes in the network; there would not be any need for the algorithm to proceed further with assessing the network-wide similarity of the nodes if the value for the Hopkins Statistic measure is low.

### 33.6 Conclusions and Future Work

Our high-level contribution in this paper is a proposal to assess the logical clusterability of the nodes in a network on the basis of node-level metrics (such as centrality metrics). We propose to distribute the nodes in a  $K$ -dimensional coordinate system (referred to as the logical topology) of the normalized centrality values of the vertices (where ' $K$ ' is the number of centrality metrics considered and we assign a dimension for each centrality metric). Two vertices that need not be physically connected in the complex network could be very close to each other in the logical topology if they incur similar values for the centrality metrics. The Hopkins Statistic has been identified as the quantitative measure to assess the similarity-based clusterability of nodes in such a logical topology. A lower value for the Hopkins Statistic is definitely an indication that the nodes do not exhibit similar values for

the centrality metrics considered. On the other hand, a higher value for the Hopkins Statistic is an indication that there could exist one or more clusters of similar vertices in the network with respect to the centrality metrics considered. We analyzed a suite of 47 real-world networks of diverse degree distributions and determined their Hopkins Statistic values with respect to the neighborhood-based degree and eigenvector centrality metrics and the shortest path-based betweenness and closeness centrality metrics. We observed about 60% of the real-world networks to exhibit relatively larger Hopkins Statistic values with respect to the neighborhood-based centrality metrics. Also, the Hopkins Statistic measure values for the neighborhood-based centrality metrics are relatively closer to 1 compared to the values for the shortest path-based centrality metrics. As part of future work, we plan to develop a network-wide node similarity index (NSI) measure for complex network analysis and test our hypothesis that a higher Hopkins Statistic is a prerequisite for a network to exhibit a higher value for the NSI.

**Acknowledgements** This research was funded by the NASA EPSCoR subaward #: NNX14AN38A from the University of Mississippi; the NSF MRI Grant 13-38192 and NSF CNS Grant 14-56638.

## References

1. Grando, F., Noble, D., Lamb, L.C.: An analysis of centrality measures for complex and social networks. In: Proceedings of IEEE Global Communications Conference, pp. 1–6, Washington, DC, USA (Dec 2016)
2. Meghanathan, N.: On the use of centrality measures to determine connected dominating sets for mobile ad hoc networks. *Int. J. Ad Hoc Ubiquitous Comput.* **26**(4), 205–221 (2017)
3. Ozgur, A., Vu, T., Erkan, G., Radev, D.R.: Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics.* **24**(13), 277–285 (2008)
4. Ding, Y.: Scientific collaboration and endorsement: network analysis of coauthorship and citation networks. *J. Informet.* **5**(1), 187–203 (2011)
5. Eslami, M., Zheng, G., Eramian, H., Levchuk, G.: Anomaly detection on bipartite graphs for cyber situational awareness and threat detection. In: Proceedings of the 2017 IEEE International Conference on Big Data, pp. 4741–4743. Boston, MA, USA (Dec 2017)
6. Newman, M.E.J.: *Networks: An Introduction*. Oxford University Press, Oxford (2010)
7. Bonacich, P.: Power and centrality: a family of measures. *Am. J. Sociol.* **92**(5), 1170–1182 (1987)
8. Freeman, L.: A set of measures of centrality based on betweenness. *Sociometry.* **40**(1), 35–41 (1977)
9. Freeman, L.: Centrality in social networks: conceptual clarification. *Soc. Networks.* **1**(3), 215–239 (1979)
10. Huang, J., Sun, H., Han, J., Deng, H., Sun, Y., Liu, Y.: SHRINK: a structural clustering algorithm for detecting hierarchical communities in networks. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, pp. 219–228, Toronto, Canada (Oct 2010)
11. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U. S. A.* **99**(12), 7821–7826 (June 2002)
12. Banerjee, A., Dave, R.: Validating clusters using the hopkins statistic. In: Proceedings of the IEEE International Conference on Fuzzy Systems, pp. 149–153, Budapest, Hungary (July 2004)
13. Szabo, G., Alava, M., Kertesz, J.: Clustering in complex networks. *Complex Networks, Lect. Notes Phys.* **650**(1), 139–162 (2004)
14. Soler, J., Tence, F., Gaubert, L., Buche, C.: Data clustering and similarity. In: Proceedings of the 26th International Florida Artificial Intelligence Research Society Conference, pp. 492–495, St. Pete Beach, FL, USA (May 2013)
15. Aalam, P., Siddique, T.: Comparative study of data mining tools used for clustering. In: Proceedings of the 3rd International Conference on Computing for Sustainable Global Development, pp. 3971–3975, New Delhi, India (Mar 2016)
16. Meghanathan, N.: Randomness index for complex network analysis. *Soc. Netw. Anal. Min.* **7**(25), 1–15 (2017)
17. Meghanathan, N.: Spectral radius as a measure of variation in node degree for complex network graphs. In: Proceedings of the 3rd International Conference on Digital Contents and Applications, pp. 30–33, Hainan, China (Dec 2014)
18. Taghva, K., Veni, R.: Effects of similarity metrics on document clustering. In: Proceedings of the 7th International Conference on Information Technology: New Generations, pp. 222–226, Las Vegas, NV (2010)
19. Han, J., Kamber, M., Pei, J.: *Data Mining: Concepts and Techniques*, 3rd edn. Morgan and Kaufmann Publishers, Burlington, MA (July 2011)
20. Blooma, M.J., Chua, A.Y.K., Goh, D.H.: Quadripartite graph-based clustering of questions. In: Proceedings of the 8th International Conference on Information Technology: New Generations, pp. 591–596, Las Vegas, NV (2011)
21. Jeh, G., Widom, J.: SimRank: a measure of structural-context similarity. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 538–543, Edmonton, Alberta, Canada (July 2002)
22. Yu, W., Lin, X., Zhang, W., Chang, L., Pei, J.: More is simpler: effectively and efficiently assessing node-pair similarities based on hyperlinks. *Proc. VLDB Endowment.* **7**(1), 13–24 (Sept 2013)