Computer and Information Science; Vol. 18, No. 1; 2025
ISSN 1913-8989 E-ISSN 1913-8997
Published by Canadian Center of Science and Education

Core-Periphery Analysis Using Principal Components of the Neighborhood-based Bridge Node Centrality Tuple

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. E-mail: natarajan.meghanathan@jsums.edu

Research was sponsored through a sub contract received from University of Virginia titled: Global Pervasive Computational Epidemiology, with the National Science Foundation as the primary funding agency.

Received: December 10, 2024 Accepted: January 10, 2025 Online Published: February 20, 2024

doi:10.5539/cis.v18n1p62 URL: https://doi.org/10.5539/cis.v18n1p62

Abstract

The neighborhood-based bridge node centrality (NBNC) tuple has been proposed in the literature to rank nodes for the extent they could serve as a bridge node. The NBNC tuple of a node v has three entries: (# components in NG_v, 1-algebraic connectivity ratio of NG_v and degree of node v), where NG_v is the neighborhood graph of node v. The research presented in this paper conducts principal component analysis on dataset comprising of NBNC tuples of all the nodes and computes a weighted PC_NBNC score based on the entries for the nodes in the dominating principal components (variances $\geq$ 1.0). The proposed model is to classify nodes as core (or peripheral) if their weighted PC_NBNC score is $\geq$ 0.0 (or < 0). The study measures the fractions of core-core, core-peripheral and peripheral-peripheral links and the fractions of core and peripheral nodes and uses these measures to classify a real-world network as either core-heavy or peripheral-heavy. Accordingly, 48 of the 80 real-world networks are classified as core-heavy (observed to be dominated by core nodes and core-core links) and the remaining 32 networks are classified as peripheral-heavy (observed to be dominated by peripheral nodes).

Keywords: Core Nodes, Peripheral Nodes, Neighborhood-based Bridge Node Centrality Tuple, Bridge Nodes, Algebraic Connectivity, Components, Node Degree

1. Introduction

Core-Periphery analysis is a mesoscopic structural analysis (Yanchenko & Sengupta, 2023) of the network aimed at categorizing nodes as core or peripheral nodes. Core nodes are considered to form a densely connected backbone forming the center of the network, whereas, the peripheral nodes are sparsely connected and are present away from the network center. There are several approaches (such as those based on spectral decomposition (Cucuringu et al., 2016), random walking (Della Rossa et al., 2013), motif counting (Ma et al., 2018), statistical inference (Kojakkku & Masuda, 2017)) in the literature to conduct core-periphery analysis and identify the core nodes and peripheral nodes. However, to the best of our knowledge, the study has not come across a bridge node-based approach to classify nodes as core nodes and peripheral nodes. Throughout the paper, the terms 'vertex' and 'node', 'link' and 'edge', 'graph' and 'network' are used interchangeably. They mean the same.

The research considers a node as a bridge node (Meghanathan, 2021) if its removal would disconnect its otherwise connected neighbors or at least make them sparingly connected. In a recent work, Meghanathan (2021) proposed a neighborhood graph-based approach to identify bridge nodes. The neighborhood graph of a node v (NG_v) comprises of the neighbors of node v and the edges connecting them. Note that NG_v does not include node v. Hence, if NG_v is not connected, it implies the neighbor nodes of v would not be connected without node v. Our hypothesis in this paper is that such bridge nodes (whose presence is critical for the connectivity of its neighbors) are candidates for being considered as core nodes of the network.

Meghanathan (2021) proposed the notion of NBNC (neighborhood-based bridge node centrality) tuple to capture the extent to which a node could play the role of a bridge node. The NBNC tuple of a node v comprises of three entries, in this order: (the number of components in NG_v, 1 - algebraic connectivity ratio of NG_v and degree of node *v*). The algebraic connectivity (ALGC; Maia de Abreu, 2007) of NG_v is the second Eigenvalue of the Laplacian matrix (Maia de Abreu, 2007) of NG_v (more details are in Section 2) and the algebraic connectivity ratio of NG_v is the ratio of the ALGC(NG_v) and the degree of node *v*. If NG_v of a node *v* is not connected (that is, NG_v has two or more components), then ALGC(NG_v) would be 0 and 1-the algebraic connectivity ratio of NG_v would be 1.0. A ranking criteria was also proposed by Meghanathan (2021) to rank nodes (for the extent they play the role of bridge nodes) per the entries of the NBNC tuple. Node *i* is ranked higher than node *j* if the number of components in NG_i is greater than the number of components in NG_j. If the number of components in NG_i and NG_j are the same, then the node that has the relatively larger value for 1-ALGC ratio is ranked higher. If two nodes incur the same values for the first two entries of the NBNC tuple, the node with the relatively larger degree is ranked higher. Two nodes are ranked equal (as bridge nodes) if all the three entries in their NBNC tuples are the same.

A closer look at the above NBNC-based ranking criteria reveals that the three entries in the NBNC tuple are likely to exhibit at least a moderately positive correlation with respect to each other. A node of larger degree is more likely to incur two or more components in its neighborhood graph compared to a node of lower degree. Also, if the neighborhood graph of a node is not connected (that is, if the first entry: number of components in the neighborhood graph is greater than 1), then the second entry (1-algebraic connectivity ratio) would take up the maximum possible value of 1.0 (as the ALGC value for a disconnected neighborhood graph would be 0.0). On the other hand, if the first entry in the NBNC tuple is 1 for a node (that is, the neighborhood graph of the node is connected), then the second entry would be less than 1.0 (as the ALGC value for a connected neighborhood graph would be greater than 0.0). Prior research (Meghanathan, 2017) has also indicated that real-world networks exhibit a negative relationship between node degree and local clustering coefficient (LCC; Newman, 2010). The LCC of a node is the probability that any two neighbors of the node are connected. For non-random networks, a high-degree node is expected to incur a lower LCC compared to a low-degree node and vice-versa (Meghanathan, 2017). This implies, the neighborhood graph of a high-degree node is more likely to comprise of two or more components (as it is less likely that all the neighbor nodes of a high-degree node would be reachable to each other) compared to the neighborhood graph of a low-degree node. Hence, the second entry of the NBNC tuple is more likely to be 1.0 for high-degree nodes and less than 1.0 for low-degree nodes.

The above inferences for at least a moderately positive correlation expected between any two entries in the NBNC tuple facilitates to conduct principal component analysis (PCA; Jolliffe, 2002) on a dataset comprising of these three entries for all the nodes in a network. The research expects just one dominating principal component (whose variance will be greater than 1.0) among the three principal components, owing to at least moderately positive correlation expected between any two entries in the NBNC tuple. The research proposes that entries for the nodes in the dominating principal component could be considered as a scalar centrality metric (referred to as PC_NBNC) that captures the extent to which a node plays the role of a bridge node. In case there are two or more dominating principal components, the PC_NBNC values for the nodes could be computed as the weighted average of the corresponding entries for the nodes in the dominating principal components (with the weights being the variances of the nodes). The research claims that highly ranked nodes (that is, high-degree nodes whose neighborhood graph is more likely to comprise of two or more components) per the NBNC tuple are more likely to incur a larger PC_NBNC value and vice-versa. Likewise, low-ranked nodes (that is, low-degree nodes whose neighborhood graph has just one component and a lower 1-ALGC ratio) are expected to incur lower, negative PC_NBNC values.

For core-periphery analysis, the research proposes that nodes with a positive PC_NBNC value (≥ 0.0) be considered as the core nodes and nodes with a negative PC_NBNC value (< 0.0) be considered as the peripheral nodes. Based on such a classification, the study determines the fractions of links between any two core nodes, between core and peripheral nodes and between any two peripheral nodes. Using these fractions of links and nodes, the study proposes a classification criteria to classify the network as either core-heavy or peripheral-heavy. If the fraction of core-core links is greater than 0.50, the study classifies the network as core-heavy. If the fraction of peripheral-peripheral links is greater than 0.50, the study classifies the network as peripheral-heavy. If either of the above two criterion are not met: if the fraction of the core nodes is greater than or equal to 0.50, the study classifies the network as core-heavy; otherwise, the study classifies the network as peripheral-heavy.

The rest of the paper is organized as follows: Section 2 reviews related work and highlights the contributions and uniqueness of our approach in this paper. Section 3 presents the computation of the NBNC tuples of the nodes in a toy example graph and illustrates the execution of PCA on a dataset built based on these NBNC tuples. Section 3 also presents the classification of the nodes in the toy example graph as core or peripheral based on their PC_NBNC values as well as the classification of the graph as core-heavy or peripheral-heavy based on the fractions of links and nodes. Section 4 presents the PC_NBNC-based results for the fractions of core-core, core-peripheral and peripheral-peripheral links and fractions of core nodes and peripheral nodes for a suite of 80 real-world networks, and accordingly presents the classification of each of these networks as core-heavy or peripheral-heavy. Section 4 also analyzes the impact of the spectral radius ratio for node degree and degree-based assortativity index on the classification of networks as core-heavy or peripheral-heavy per the PC_NBNC values. Section 5 concludes the paper and presents plans for future work.

2. Related Work

This section will review some of the quantitative metrics and properties proposed in the literature related to core-peripheral analysis and how PC_NBNC compares to them and satisfies the properties. Per Borgatti & Everett (2000) and Zhang et al (2015), core nodes are high-degree nodes, but a high-degree node need not necessarily be a core node. The formulation of the PC_NBNC scalar centrality metric confirms the above statement. PC_NBNC takes into consideration the degree of the nodes; but also gives importance to the connectivity among the neighbors of a node, which is not entirely dependent on the degree of the node. The covariance matrix shown in Figure 4 (in the case of the toy example graph used in Section 2) for the three features (# components in the neighborhood graph NG_v of a node *v*, 1-Algebraic connectivity(NG_v), and degree of node *v*) used in the formulation of the PC_NBNC is a testament to our claim that beyond degree, there are other characteristics that contribute to the role played by a node as a core node. In Figure 4, the Pearson's correlation coefficients between DEG(node *v*) and the other two features are 0.4588 and 0.7189.

k-core decomposition (Newman, 2010) is a classical method used to extract the densely connected sub graph(s) of a graph, wherein each constituent node have at least *k*-neighbors that are also present in the same sub graph. The procedure to obtain the *k*-core of a graph for a given *k* is as follows: First remove all the vertices with degree less than *k*. This would lead to the degree of some other vertices becoming less than *k*. Then, remove such vertices as well and continue the process until there are one or more sub graph(s) in which each constituent node has at least *k* neighbors present in the same sub graph. The *k-core index of a node v* is the largest value of *k* for which node *v* would be part of a *k*-core sub graph and not part of a *k*'-core sub graph (wherein $k' > k$).

While proposing a stochastic block model for determining core-peripheral structures in complex networks, Gallgher et al (2021) proposed the following inequality to expect among the probabilities p_{cc} (probability for a link between two core nodes), p_{cp} (probability for a link between a core node and a peripheral nodes) and p_{pp} (probability for a link between two peripheral nodes): $p_{cc} > p_{cp} > p_{pp}$. However, the study claims that this inequality need not be necessarily satisfied by the core-peripheral structures for all complex real-world networks (as is also observed in this research; see Figure 13 in Section 3).

Several works (for example: Kojakku & Masuda, 2017 and Puck Rombach et al., 2014) in the literature have advocated the presence of a multi-core and/or multi-periphery architecture rather than a single hub-and-spoke architecture. Barbera et al (2015) observed the non-core nodes to play a crucial role in the effective spread of social media protests from the core nodes (forming the epicenter) to the rest of the nodes in the network. Though the non-core nodes may not have as many neighbors as the core nodes, the sheer presence of a large number of non-core nodes around the core nodes were observed to be critical for information diffusion. This observation was also earlier made by Kirsak et al (2010) who noted that the influential spreaders in a social network need not necessarily be the most central nodes with high degree centrality or betweenness centrality. The numerical values for the PC_NBNC metric observed for the toy example graph as well as for the real-world networks also confirm the possible presence of a multi-core multi-periphery architecture in complex networks. As of now, the study classifies that all nodes incurring a positive PC_NBNC value be referred to core nodes and those with negative PC_NBNC values as peripheral nodes. However, note that the PC_NBNC values of the nodes do significantly differ among the core nodes as well as among the peripheral nodes. Consider the PC_NBNC values shown in Figure 4 for the nodes in the toy example graph. The PC_NBNC values for nodes 1, 2, 5 and 7 classified as core nodes are respectively {2.8547, 1.5997, 0.6082, 0.7537} and those for nodes 3, 4, 6, 8, 9 and 10 classified as peripheral nodes are respectively {-1.0722, -0.1362, -1.0722, -0.1362, -1.6997, -1.6997}. Such diverse PC_NBNC values for the core nodes and peripheral nodes clearly advocate for a multi-core/multi-peripheral architecture as well as facilitate to be able to determine the same. For example, per the above PC_NBNC values for the nodes in the toy example graph: there could be three layers of core nodes {1}, {2}, {5, 7}, say from the inner core to the outer core; and likewise there could be three layers of peripheral nodes {4, 8}, {3, 6}, {9, 10} from the inner periphery to the outer periphery. As part of future work, plan to run clustering algorithms such as K-Means (Lloyd, 1982) and DBSCAN (Hashler et al., 2019) to extract clusters of core nodes based on their PC_NBNC values.

3. NBNC-based Core-Periphery Analysis for a Toy Example Graph

The section first presents the notions of the neighborhood graph of a node and the Laplacian matrix of a graph (also applicable to the neighborhood graphs) that are used in the NBNC calculations. The neighborhood graph of a node *v* (denoted: NG_v) comprises of the neighbors of node *v* as the vertices and the edges connecting these neighbors. Note that NG_v does not include node *v*. The Laplacian matrix of a graph comprises one of the following values for entries (*i*, *j*): if *i* equals *j* (diagonal entries), the value of the entry is the degree of node *i* otherwise (*i* is not equal to *j*), then the value of the entry is either a 0 (if there is no edge between vertices *i* and *j*) or -1 (if there is an edge between vertices *i* and *j*).

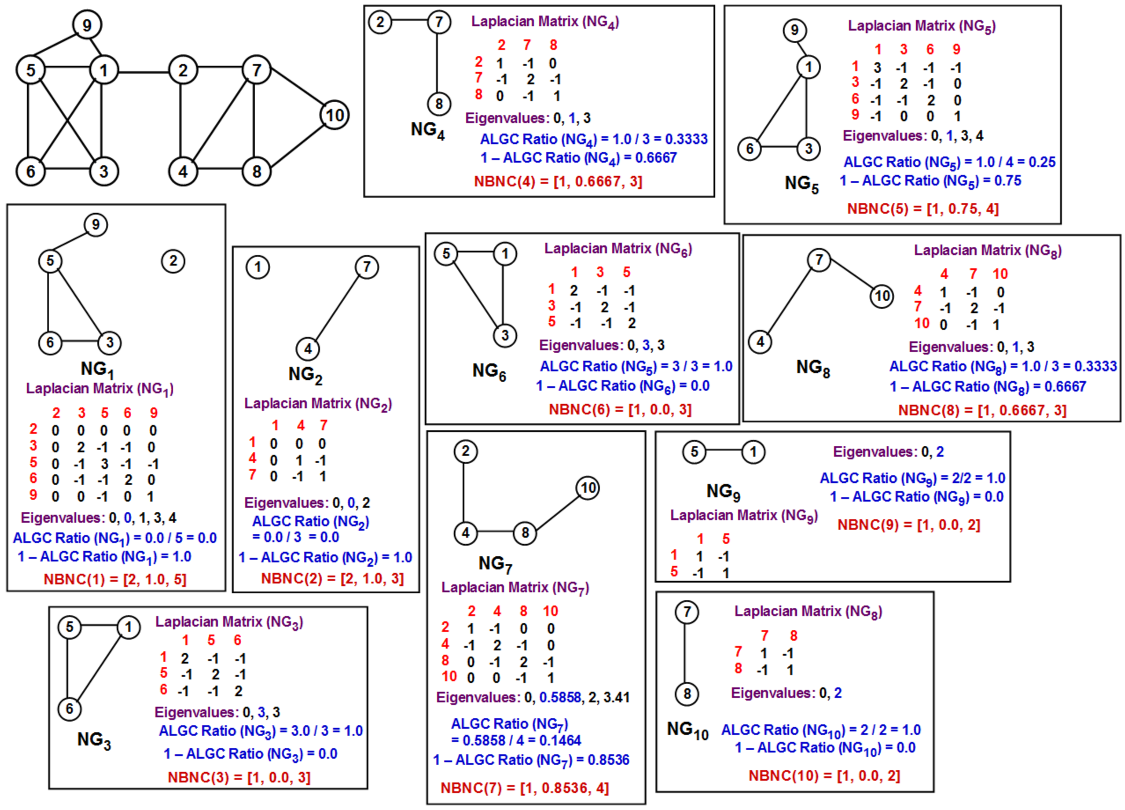


Figure 1. NBNC Calculations for the Nodes in a Toy Example Graph

The Eigenvalues of the Laplacian matrix are greater than or equal to 0.0 (Maia de Abreu, 2007). The number of 0s among the Eigenvalues of the Laplacian matrix correspond to the number of components in the underlying graph represented by the Laplacian matrix (Maia de Abreu, 2007). Hence, if the underlying graph represented by the Laplacian matrix is a connected graph (that is, all the vertices of the graph are in the same component), then there would be only one 0 among the Eigenvalues of the matrix. On the other hand, if the underlying graph represented by the Laplacian matrix is not connected, the number of 0s among the Eigenvalues would be greater than 1, with the number of 0s corresponding to the number of components.

Figure 1 presents the toy example graph of 10 nodes used to illustrate NBNC-based core-periphery analysis. Figure 1 also presents the neighborhood graphs for each node in the toy example graph and their NBNC calculations. The algebraic connectivity (ALGC; Maia de Abreu, 2007) of a neighborhood graph is the second Eigenvalue of the Laplacian matrix of the neighborhood graph. Figure 1 presents the Laplacian matrices for each of the neighborhood graphs and their corresponding Eigenvalues. The ALGC ratio for a neighborhood graph is the ratio of the ALGC value for the neighborhood graph divided by the number of nodes in the neighborhood graph. The NBNC tuple for a node *v* is [# components in NG_v, 1-ALGC ratio(NG_v), degree of node *v*].

Figure 2 presents the NBNC tupes for the nodes (in the order of the ranking of the nodes per the entries in the NBNC tuple) in the toy example graph of Figure 1. The tentative rank ids are unique numbers assigned for the nodes per their ranking criteria: the higher the rank, the lower the tentative rank id and vice-versa. If two or more nodes incur the same values for all the three entries in the NBNC tuple, the tie is broken in favor of their node ids (that is, the node with the lower id gets the lower tentative rank). The final rank ids reflect the actual ranking of the nodes with respect to their role as bridge nodes. If a node does not encounter tie with any other node, its final rank id is the same as its tentative rank id. If two or more nodes encounter tie per their NBNC tuple entries, then their final rank id is the average of their tentative rank ids.

Node ID	NBNC Tuple	Tentative Rank	Final Rank
1	[2, 1.0, 5]	1	1
2	[2, 1.0, 3]	2	2
7	[1, 0.8536, 4]	3	3
5	[1, 0.75, 4]	4	4
4	[1, 0.6667, 3]	5	5.5
8	[1, 0.6667, 3]	6	5.5
3	[1, 0.0, 3]	7	7.5
6	[1, 0.0, 3]	8	7.5
9	[1, 0.0, 2]	9	9.5
10	[1, 0.0, 2]	10	9.5

Figure 2. NBNC-based Ranking of the Nodes in a Toy Example Graph

Figure 3 presents the dataset comprising of the 10 nodes (as record ids) and the three entries (# components in NG_v, 1-ALGC ratio(NG_v) and DEG(v), for a node v) of the NBNC tuples of the nodes as the features. Figure 3 presents both the raw dataset and its standardized version (determined using the average and standard deviation of the feature values in the raw dataset). Figure 4 presents the results of PCA on this dataset. PCA comprises of two steps: (1) Covariance matrix computation and (2) Eigenvalue decomposition. The entries in the Covariance matrix reflect the correlation (Pearson's correlation coefficient; Strang, 2016) between any two features of the dataset. The study then determines the Eigenvalues and their corresponding Eigenvectors for the Covariance matrix. The entries in the Eigenvectors reflect the loadings for the three features. One then multiplies the standardized dataset with each of the three Eigenvectors to obtain the three principal components (PC) of the dataset. The study observes PC-1 to be the only PC with a variance (2.1957) greater than or equal to 1.0 and is thereby considered the only dominating PC. Note that the sum of the variances of the PCs should be 3, corresponding to the number of features considered in the dataset. The dominating PC, PC-1, captures 2.1957/3 ~ about 73% of the variation in the dataset.

Node ID	# Comp (NG_V)	1-ALGC Ratio(NG_V)	DEG(v)
1	2	1	5
2	2	1	3
3	1	0	3
4	1	0.6667	3
5	1	0.75	4
6	1	0	3
7	1	0.8536	4
8	1	0.6667	3
9	1	0	2
10	1	0	2

Node ID	# Comp (NG_V)	1-ALGC Ratio(NG_V)	DEG(v)
1	1.8975	1.1507	1.9589
2	1.8975	1.1507	-0.2177
3	-0.4744	-1.1220	-0.2177
4	-0.4744	0.3932	-0.2177
5	-0.4744	0.5825	0.8706
6	-0.4744	-1.1220	-0.2177
7	-0.4744	0.8180	0.8706
8	-0.4744	0.3932	-0.2177
9	-0.4744	-1.1220	-1.3059
10	-0.4744	-1.1220	-1.3059

Figure 3. NBNC-based Raw Dataset and its Standardized Version used for PCA

Since there is only one dominating PC (PC-1), the study considers the entries for the nodes in PC-1 as a scalar quantification (referred to as the PC_NBNC metric) of the extent with which the nodes play the role of bridge nodes per the NBNC tuple. If there were more than one dominating PCs, the PC_NBNC metric for a node is computed as the weighted average of the entries for the node in the dominating PCs, with the variances of the dominating PCs used as weights. Figure 4 also presents the classification of the nodes as core [C] or peripheral [P] nodes based on their PC_NBNC values: if the PC_NBNC value for a node is positive (≥ 0.0), the node is classified as core; otherwise (that is, negative, < 0), the node is classified as peripheral. Per this classification rule, the study observes nodes 1, 2, 5 and 7 to be classified as core nodes and the rest of the nodes (nodes 3, 4, 6, 8, 9 and 10) classified as peripheral nodes. Figure 4 presents the rankings of the nodes based on their PC_NBNC scores. The study observes the rankings (including the ties) to be the same per both the NBNC tuple (see Figure

2) and the scalar metric PC_NBNC (see Figure 4).

	# Comp (NG_v)	1-ALGC Ratio(NG_v)	DEG(v)
# Comp (NG_v)	1.0000	0.6064	0.4588
1-ALGC Ratio(NG_v)	0.6064	1.0000	0.7189
DEG(v)	0.4588	0.7189	1.0000

Covariance Matrix of the '3' Features

Eigenvalues	19.76	4.98	2.26
	Eigenvectors, EV		
	EV-1	EV-2	EV-3
# Comp (NG_v)	0.5346	-0.7995	-0.2740
1-ALGC Ratio(NG_v)	0.6178	0.1485	0.7722
DEG(v)	0.5766	0.5821	-0.5733

Eigenvalue Decomposition of the Covariance Matrix

Node ID	PC-1	PC-2	PC-3
1	2.8547	-0.2058	-0.7544
2	1.5997	-1.4726	0.4934
3	-1.0722	0.0859	-0.6116
4	-0.1362	0.3109	0.5583
5	0.6082	0.9725	0.0806
6	-1.0722	0.0859	-0.6116
7	0.7537	1.0074	0.2624
8	-0.1362	0.3109	0.5583
9	-1.6997	-0.5475	0.0123
10	-1.6997	-0.5475	0.0123
Variance	2.1957	0.5533	0.2511

Principal Components and their Variances

Node ID	PC_NBNC	Core [C], Peripheral [P] Classification	Tentative Rank	Final Rank
1	2.8547	C	1	1
2	1.5997	C	2	2
3	-1.0722	P	7	7.5
4	-0.1362	P	5	5.5
5	0.6082	C	4	4
6	-1.0722	P	8	7.5
7	0.7537	C	3	3
8	-0.1362	P	6	5.5
9	-1.6997	P	9	9.5
10	-1.6997	P	10	9.5

Classification and Rankings of the Nodes per the PC_NBNC Metric

Figure 4. Principal Component Analysis of the NBNC-based Dataset for Core-Periphery Analysis

Figure 5 presents a visualization of the toy example graph on the basis of the core/peripheral classification per the PC_NBNC values. Figure 5 also presents the classification of the links as core-core [C..C], core-peripheral [C..P] and peripheral-peripheral [P..P] links, for a total of 16 links. The fractions of core nodes and peripheral nodes are respectively 4/10 = 0.40 and 6/10 = 0.60. The fractions of C..C, C..P and P..P links are respectively: 3/16 = 0.1875, 10/16 = 0.625 and 3/16 = 0.1875. Since the fraction of C..P links is greater than 0.50, the study uses the fractions of core nodes and peripheral nodes to decide the classification as core-heavy or peripheral-heavy. Since the fraction of peripheral nodes is greater than 0.50, the study classifies the network as peripheral-heavy.

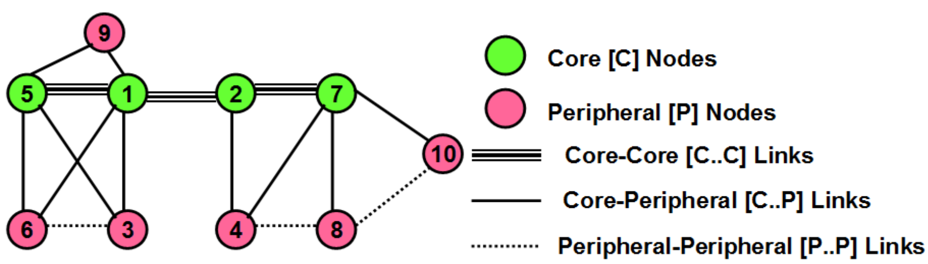


Figure 5. PC_NBNC-based Core-Peripheral Classification of the Nodes and Links in the Example Graph

Figure 6 formally presents the classification criteria for core-heavy network or peripheral-heavy network. Per these rules, it is obvious that the study gives preference for fractions of links (rather than fractions of nodes) to classify a network as core-heavy or peripheral-heavy. This is because (since PC_NBNC is based on PCA), the sum of the PC_NBNC values for all the nodes is expected to be 0.0 and for such a property to hold good, a reasonable fraction of nodes should incur negative PC_NBNC values (that is, be classified as peripheral nodes) as well as a reasonable fraction of nodes should incur positive PC_NBNC values (that is, be classified as core nodes); the study expects the fractions of core nodes to be very much comparable to the fractions of peripheral nodes. Nevertheless, the presence or absence of links between the core nodes, between the core nodes and peripheral nodes and between the peripheral nodes would be a more appropriate reflection on the classification of the network as core-heavy or peripheral-heavy. The core nodes are expected to be connected to other core nodes as well as to one or more peripheral nodes, whereas the peripheral nodes are connected to one or few core nodes, but not expected to be connected to other peripheral nodes. Nevertheless, the fraction of the core-core links, core-peripheral links and peripheral-peripheral links is likely to vary among networks.

(i) If the fraction of core-core links is greater than 0.50:

Classify a network to be core-heavy.

(ii) If the fraction of peripheral-peripheral links is greater than 0.50:

Classify a network to be peripheral-heavy.

(iii) If the fraction of core-peripheral links is greater than 0.50 (or) if all the three fractions of links are each less than 0.50 :

if the fraction of core nodes is 0.50 or above:

Classify a network to be core-heavy

else // that is, if the fraction of peripheral nodes is greater than 0.50:

Classify a network to be peripheral-heavy

Figure 6. Classification Criteria for Core-heavy or Peripheral-heavy Network based on the Fractions of Core-Core Links, Core-Peripheral Links, Peripheral-Peripheral Links and Fractions of Core Nodes and Peripheral Nodes

4. NBNC-based Core-Periphery Analysis for Real-World Networks

This section presents the results of the PC_NBNC scores-based core-periphery analysis for a suite of 80 real-world networks spanning diverse network domains (listed alongside the network names in Figures 7 and 8) and spectral radius ratio of node degree (λ_{sp}^{k} ; Meghanathan 2014) ranging from 1.01 to 7.53. Figures 7 and 8 present the results that include the following: fractions of core-core links ($f_{links}^{C..C}$), core-peripheral links ($f_{links}^{C..P}$), peripheral-peripheral links ($f_{links}^{P..P}$) and the fractions of core nodes (f_{nodes}^{C}) and peripheral nodes (f_{nodes}^{P}) as well as the classification of the network as core-heavy (denoted by a C) or peripheral-heavy (denoted by a P) per the classification criteria listed in Figure 6.

In Figures 7 and 8, the cells are highlighted for which the fraction of links and/or nodes were greater than or equal to 0.50 and thereby enabling to classify (as either core-heavy or peripheral-heavy) the real-world networks. Figure 9-(i) presents the distribution of the three fractions of links: an important observation to make is that the fraction of core-core links exhibit a trend to decrease with increase in the fraction of core-peripheral links, whereas the fraction of peripheral-peripheral links exhibit a trend to increase with increase in the fraction of core-peripheral links. Such a link distribution is justified by the distribution of the fractions of core nodes and peripheral nodes vis-a-vis the fraction of core-peripheral links; see Figure 9-(ii). With increase in the fraction of core-peripheral links, the study observes a decrease in the fraction of core nodes and an increase in the fraction of peripheral nodes.

Of the 80 real-world networks, the study observes 48 networks to be core-heavy and 32 networks to be peripheral-heavy. The fraction of core-core links were found to be greater than or equal to 0.50 for 35 of the 80 real-world networks, whereas the fraction of peripheral-peripheral links were found to be greater than or equal to 0.50 for only 2 of the 80 real-world networks. This also implies 30 of the 32 real-world networks that were classified as peripheral-heavy attained such a classification due to the fraction of peripheral nodes being greater than 0.50. On the other hand, for 35 of the 48 core-heavy networks, the presence of a larger fraction of the core-core links itself was sufficient to attain such a classification; for only 13 of the 48 core-heavy networks, the presence of a large fraction of core nodes (≥ 0.50) was used as the deciding criterion for the core-peripheral classification of the networks. The fraction of core-peripheral links were observed to be greater than or equal to 0.50 for 18 of the 80 real-world networks (of these, 15 networks were classified as peripheral-heavy); none of the three fractions of links were observed to be greater than or equal to 0.50 for 25 of the 80 real-world networks (of these, 15 networks were classified as peripheral-heavy). All the above inferences indicate that the core-heavy networks attained such a classification mostly by virtue of a larger fraction of the core-core links; whereas, the peripheral-heavy networks attained such a classification mostly by virtual of a larger fraction of peripheral nodes. Note that the fractions of core nodes and peripheral nodes were only used to classify the networks as core-heavy or peripheral-heavy if the fractions of core-core links and peripheral-peripheral links could not be used (that is, were less than 0.50) for the classification.

#	Net. Name	Domain	λ_{sp}^{k}	$f_{links}^{C...C}$	$f_{links}^{C...P}$	$f_{links}^{P...P}$	f_{nodes}^{C}	f_{nodes}^{P}	C/P
1	Lazega Law Firm	Employment	2.6324	0.0146	0.9854	0.0000	0.0423	0.9577	P
2	CS Dept. Arhus Net.	Employment	2.1200	0.3059	0.6941	0.0000	0.2787	0.7213	P
3	Hepatitus C Genetic	Biological	4.1707	0.0081	0.9919	0.0000	0.0190	0.9810	P
4	CKM Physicians Net.	Social	4.7446	0.0045	0.9955	0.0000	0.0124	0.9876	P
5	World Trade Net.	Miscellaneous	1.3774	0.3486	0.6000	0.0514	0.4000	0.6000	P
6	Slov. Magazine Net.	Literature	1.0478	0.3562	0.5449	0.0990	0.5323	0.4677	C
7	Manuf. Comp. Emp.	Employment	1.1205	0.4003	0.4501	0.1496	0.5065	0.4935	C
8	Ht2009 Dyn, Conf.	Friendship	1.2082	0.4873	0.4424	0.0703	0.5310	0.4690	C
9	Fratern. College Dorm	Friendship	1.1106	0.4612	0.4550	0.0838	0.5690	0.4310	C
10	Cat Brain Network	Biological	1.1970	0.4178	0.4603	0.1219	0.4769	0.5231	P
11	Soccer World Cup'98	Miscellaneous	1.4484	0.1695	0.4492	0.3814	0.3714	0.6286	P
12	gd96Net	Literature	2.3751	0.7424	0.2576	0.0000	0.6778	0.3222	C
13	Adjacency Noun Net.	Co-appearance	1.7327	0.5176	0.4212	0.0612	0.5179	0.4821	C
14	Senator Press Meets.	Social	1.5660	0.0587	0.3354	0.6059	0.5000	0.5000	P
15	Primary School Net	Social	1.2237	0.5864	0.3311	0.0825	0.6134	0.3866	C
16	Florida Food Web Net	Biological	1.2224	0.6011	0.3319	0.0670	0.6016	0.3984	C
17	Firm High Tech Net.	Friendship	1.4400	0.6190	0.3673	0.0136	0.5455	0.4545	C
18	UK Faculty Network	Employment	1.3536	0.4055	0.4887	0.1057	0.4691	0.5309	P
19	Mexican Political Net	Social	1.2254	0.2479	0.5726	0.1795	0.4000	0.6000	P
20	Karate Network	Social	1.4659	0.5385	0.3718	0.0897	0.5000	0.5000	C
21	Dutch Literature 1976	Literature	1.4878	0.6500	0.3250	0.0250	0.6286	0.3714	C
22	Mod Math Net	Friendship	1.2566	0.5410	0.3607	0.0984	0.6333	0.3667	C
23	Prison Friendship Net	Friendship	1.3190	0.5165	0.3846	0.0989	0.5672	0.4328	C
24	Webster Ressidence	Friendship	1.2722	0.1001	0.4127	0.4872	0.4194	0.5806	P
25	Band Jazz Net	Social	1.4452	0.6131	0.3297	0.0573	0.5505	0.4495	C
26	Flying Team Cadets	Friendship	1.2103	0.3353	0.5000	0.1647	0.5208	0.4792	C
27	US States Net	Miscellaneous	1.2470	0.4953	0.4019	0.1028	0.5510	0.4490	C
28	Facebook Net 1	Social	1.8109	0.3336	0.4328	0.2335	0.5154	0.4846	C
29	Korean Family Plan.	Social	1.5252	0.0476	0.3810	0.5714	0.3714	0.6286	P
30	Facebook Net 2	Social	2.0738	0.4803	0.4004	0.1193	0.5089	0.4911	C
31	Football Network	Miscellaneous	1.0112	0.2023	0.4568	0.3409	0.4522	0.5478	P
32	Euro Road Net	Transportation	1.6613	0.2781	0.4460	0.2759	0.3169	0.6831	P
33	McCarty Social Net J	Literature	3.4808	0.3152	0.3200	0.3648	0.2337	0.7663	P
34	NetSci Collaborators	Literature	5.0682	0.3636	0.3147	0.3217	0.2512	0.7488	P
35	CS Phd Net	Employment	3.3466	0.3317	0.6683	0.0000	0.3220	0.6780	P
36	Faux Mesa High Sch.	Friendship	2.0138	0.5099	0.4307	0.0594	0.4898	0.5102	C
37	Literature Curated In.	Literature	4.2345	0.5682	0.3730	0.0588	0.4577	0.5423	C
38	Drug Network	Friendship	1.9633	0.6588	0.2967	0.0445	0.5943	0.4057	C
39	Pol. Books Network	Literature	1.4206	0.5374	0.3333	0.1293	0.6667	0.3333	C
40	Norwegian Directors	Employment	2.1014	0.1888	0.3401	0.4711	0.1893	0.8107	P

Figure 7. Fractions of Core-Core [C..C], Core-Peripheral [C..P] and Peripheral-Peripheral [P..P] Links; Fractions of Core [C] and Peripheral [P] Nodes; Classification of Real-World Networks # 1 to # 40 as Core or Peripheral based on the PC_NBNC Values of the Nodes

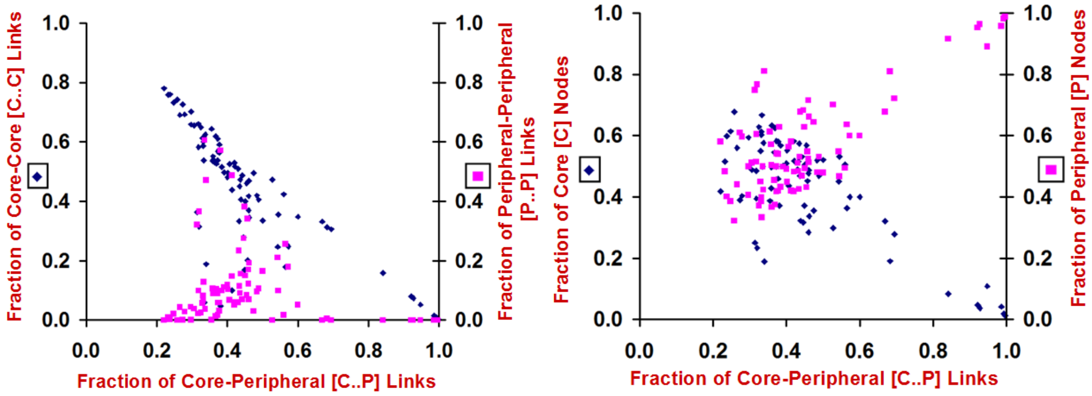
In Figure 9-i, it can be observed that the bulk of the blue data points (representing the fraction of core-core links) is in the top left, whereas the bulk of the pink data points (representing the fraction of peripheral-peripheral links) is in the bottom left. On the other hand, in Figure 9-ii, majority of the blue data points (representing the fraction of core nodes) as well as pink data points (representing the fraction of peripheral nodes) are distributed in the range of [0.40,..., 0.60]. The median of the fraction of core-core links is 0.4760, whereas the median of the fraction of peripheral-peripheral links is 0.0690; the median of the fraction of core-peripheral links is 0.4096. With 60% of the 80 real-world networks classified as core-heavy (and more than 2/3rds of the core-heavy networks attained such a classification by virtue of the dominant presence of the core-core links), all of the above numbers as well as the distribution of the data points in Figures 9-i and 9-ii are reasonable to expect among the core nodes as well as among the peripheral nodes at the core-peripheral mesoscopic level in complex networks. Overall, the study could confidently conclude that the results presented in this section justify our use of the PC_NBNC based node scores for core-peripheral analysis.

#	Net. Name	Domain	λ_{sp}^{k}	$f_{links}^{C...C}$	$f_{links}^{C...P}$	$f_{links}^{P...P}$	f_{nodes}^{C}	f_{nodes}^{P}	C/P
41	Geiser Commic Net	Social	2.5378	0.3700	0.4600	0.1700	0.2848	0.7152	P
42	Dolphin Network	Social	1.4025	0.5912	0.3774	0.0314	0.5806	0.4194	C
43	Scotland Directorate	Employment	1.8161	0.4693	0.4609	0.0698	0.4868	0.5132	P
44	Inf. Socio Patterns	Social	1.6884	0.4392	0.4127	0.1481	0.5793	0.4207	C
45	Yeast Two Hybrid PPI	Biological	4.2911	0.4057	0.4377	0.1566	0.3223	0.6777	P
46	San Juan Sur Net	Social	1.2937	0.4194	0.4839	0.0968	0.5200	0.4800	C
47	Citation GD94-2000	Literature	1.8636	0.6906	0.2656	0.0438	0.5598	0.4402	C
48	Teen. Women Friend.	Friendship	1.4031	0.4508	0.4344	0.1148	0.5745	0.4255	C
49	Strike Saw Mill Net.	Social	1.2222	0.5263	0.3684	0.1053	0.6250	0.3750	C
50	Erdos 971 Network	Literature	2.7532	0.7321	0.2473	0.0205	0.6143	0.3857	C
51	Celegans Network	Biological	1.6845	0.4972	0.3911	0.1117	0.5017	0.4983	C
52	US Airports '97 Net	Transportation	3.2195	0.6929	0.2789	0.0282	0.4036	0.5964	C
53	Les Miserable Net	Co-appearance	1.8198	0.5354	0.3583	0.1063	0.4286	0.5714	C
54	Pol Blogs Net	Literature	2.7124	0.7589	0.2327	0.0084	0.5163	0.4837	C
55	Huckleberry Coappear	Co-appearance	1.6621	0.3455	0.4618	0.1927	0.3378	0.6622	P
56	Yeast Interactome Net	Biological	4.5090	0.5821	0.3184	0.0995	0.3951	0.6049	C
57	Roget Network	Literature	1.6650	0.5515	0.3591	0.0894	0.4970	0.5030	C
58	Taro Exchange Net.	Social	1.0579	0.1795	0.5641	0.2564	0.3636	0.6364	P
59	Celegans Metabolic	Biological	2.9427	0.5254	0.4064	0.0681	0.4371	0.5629	C
60	Flensburg Bio Netw.	Biological	1.8038	0.4236	0.5593	0.0171	0.5056	0.4944	C
61	Mtuberculosis Trans.	Biological	6.1316	0.3127	0.6820	0.0053	0.1905	0.8095	P
62	Madrid Train Bombing	Friendship	1.7819	0.4733	0.4362	0.0905	0.4688	0.5313	P
63	EU Air Transport. Net	Transportation	3.8111	0.7799	0.2201	0.0000	0.4187	0.5813	C
64	RVU Email Network	Employment	2.1560	0.6554	0.3062	0.0384	0.4881	0.5119	C
65	Java Dependency Net	Literature	4.2544	0.6610	0.3172	0.0218	0.4850	0.5150	C
66	Copper Field Network	Co-appearance	1.8309	0.5640	0.3768	0.0591	0.4598	0.5402	C
67	Centrality Literature	Literature	1.8588	0.6248	0.3377	0.0375	0.5763	0.4237	C
68	Glossary Networks	Co-appearance	1.8710	0.6102	0.3729	0.0169	0.5821	0.4179	C
69	Anna Karnenina Net	Co-appearance	2.4753	0.4949	0.4746	0.0304	0.3551	0.6449	P
70	London Transport Net	Transportation	3.6004	0.4734	0.5266	0.0000	0.2992	0.7008	P
71	Mouse Transcript Net	Biological	4.2970	0.7020	0.2971	0.0008	0.5009	0.4991	C
72	Yeast Phosphoryl. Net	Biological	4.8968	0.7262	0.2733	0.0005	0.3902	0.6098	C
73	Rat Trancription Net.	Biological	3.4018	0.7601	0.2372	0.0027	0.5984	0.4016	C
74	Perl Developers Net.	Social	5.2175	0.6442	0.3548	0.0009	0.3874	0.6126	C
75	Wind Surfers Beach	Social	1.1598	0.5298	0.4196	0.0506	0.5814	0.4186	C
76	Graph Draw06 Net.	Literature	1.8027	0.0526	0.9474	0.0000	0.1089	0.8911	P
77	Macaque Dominance	Social	1.0367	0.2468	0.5424	0.2108	0.4516	0.5484	P
78	Human Herpes4 Gen.	Biological	6.0706	0.0731	0.9269	0.0000	0.0370	0.9630	P
79	Xenopus Genetic Int.	Biological	7.5333	0.1592	0.7408	0.0000	0.0846	0.9154	P
80	Gallus Genetic Int. Net	Biological	7.0001	0.0797	0.9203	0.0000	0.0479	0.9521	P

Figure 8. Fractions of Core-Core [C..C], Core-Peripheral [C..P] and Peripheral-Peripheral [P..P] Links; Fractions of Core [C] and Peripheral [P] Nodes; Classification of Real-World Networks # 41 to # 80 as Core or Peripheral based on the PC_NBNC Values of the Nodes

4.1 Impact of the Spectral Radius Ratio for Node Degree

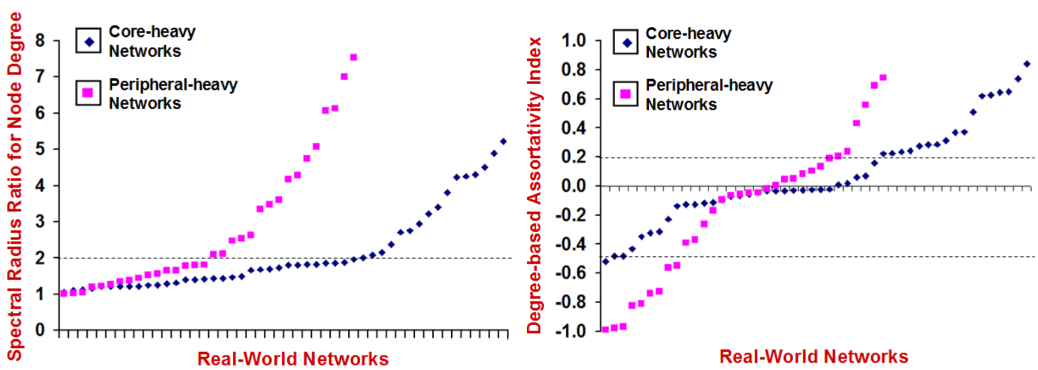
The Spectral radius ratio of node degree ($\lambda_{sp}^{k} \geq 1.0$) is a measure of the variation in node degree, independent of the number of nodes and edges in the network. The λ_{sp}^{k} values for the 80 real-world networks analyzed in this work range from 1.01 to 7.53. Figure 10-(i) presents a distribution of the 48 core-heavy and 32 peripheral-heavy networks in the sorted order of their λ_{sp}^{k} values. The median λ_{sp}^{k} value for the 48 core-heavy networks is 1.77, whereas the median λ_{sp}^{k} value for the 32 peripheral-heavy networks is 1.97. the study could use λ_{sp}^{k} = 2.0 as the cut off as 60% (48 networks) of the 80 real-world networks exhibit λ_{sp}^{k} values less than 2.0. Of these 48 real-world networks with λ_{sp}^{k} less than 2.0, only 16 networks are classified as peripheral-heavy. In other words, if a real-world network is said to be of $\lambda_{sp}^{k} < 2.0$, there is a 2/3rd chance for it to be core-heavy and only a 1/3rd chance for it to be peripheral-heavy. On the other hand, of the 32 real-world networks that incurred λ_{sp}^{k} values greater than 2.0, an equal number of networks (16 each) were observed to be core-heavy or peripheral-heavy. In other words, if a real-world network is said to be of $\lambda_{sp}^{k} > 2.0$, there is an equal chance for it to be either core-heavy or peripheral-heavy. Also, with respect to the number of dominating principal components (PCs) observed for the NBNC datasets of the 80 real-world networks, only 12 real-world networks are observed to incur more than one dominating PC. It is interesting to observe that all these 12 real-world networks were those with λ_{sp}^{k} values less than 2.0, with a median of 1.44. Of these 12 real-world networks, seven networks were classified as core-heavy and five networks were classified as peripheral-heavy.



(i) Distribution of the Fractions of Core-Core and Peripheral-Peripheral Links

(ii) Distribution of the Fractions of Core Nodes and Peripheral Nodes

Figure 9. Distribution of the Fraction of Core-Peripheral Links vs. the Link Fractions and Node Fractions



(i) Spectral Radius Ratio for Node Degree

(ii) Degree-based Assortativity Index

Figure 10. Real-World Networks Plotted in the Increasing Order of their Spectral Radius Ratio for Node Degree and Assortativity

4.2 Impact of the Degree-based Assortativity Index

The assortativity index (AssI; Newman, 2010) of a network is a quantitative measure of the extent of similarity of the end vertices of the edges with respect to a node-level metric. The AssI value for a network with respect to a particular node-level metric (Meghanathan, 2024) is computed as the Pearson's correlation coefficient of the node-level metric values for the end vertices of the edges. The AssI values range from -1 to 1; networks with AssI values less than 0 and greater than 0 are respectively considered to be dissortative and assortative with respect to the particular node-level metric used in the assortativity calculations. The degree centrality of the nodes is the often used metric to assess the assortativity of real-world networks (Meghanathan 2016). Figure 10-ii presents a distribution of the 48 core-heavy networks and 32 peripheral-heavy networks in the increasing order of their degree-based AssI (referred to as DEG_AssI) values. It is observed that all the nine real-world networks with DEG_AssI values less than -0.50 to be peripheral-heavy; whereas, 27 of the 32 real-world networks with DEG_AssI values greater than 0.20 were observed to be assortative. On the other hand, of the 49 real-world networks whose DEG_AssI values were in the range of [-0.50, ..., 0.20], only 18 real-world networks (that is, ~ 36% of the real-world networks) were observed to be peripheral-heavy. The study could thus use

DEG_AssI values of -0.50 and 0.20 as cutoffs to come to the following conclusions. If a real-world network incurred a DEG_AssI value less than -0.50, it is almost guaranteed to be peripheral-heavy; whereas, if a real-world network incurred a DEG_AssI value greater than 0.20, it is very likely (with a probability of 27/32 ~ 0.85) to be core-heavy. On the other hand, if the DEG_AssI value for a real-world networks falls in the range of [-0.50, ..., 0.20], there is a 2/3rd chance for it to be core-heavy and only a 1/3rd chance for it to be peripheral-heavy.

4.3 Impact of the Network Domain

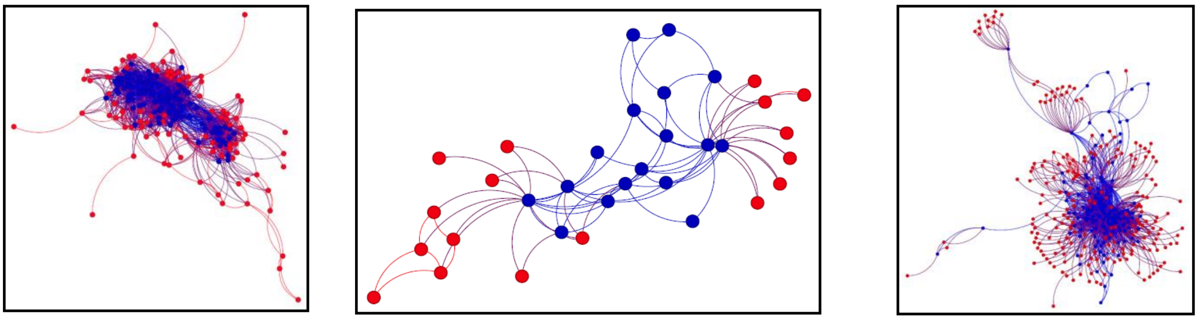
Seventy six of the eighty real-world networks analyzed in this research are observed to span 7 domains. Table 1 presents the domain names, the network #s (of Figures 7 and 8) that fall under these domains and the number (count) of these networks that are classified as core-heavy or peripheral-heavy. The network #s that are core-heavy are highlighted in blue and the network #s that are peripheral-heavy are highlighted in red. The Friendship networks and Literature networks are observed to be predominantly core-heavy: implying that friends of people who have more friends also have more friends and also serve as bridge nodes to diverse communities of friends; likewise, authors prefer to co-author with researchers who themselves have co-authored with several researchers as well as serve as bridge nodes in bringing together researchers of diverse expertise and domains. The Employment networks were predominantly peripheral-heavy, justifying the hierarchy in organizations and the connections among people follow this hierarchy. The other network domains (Biological, Social, Co-appearance and Transportation networks) appear to comprise of a comparable number of networks from both the categories. An interesting observation is that of the four transportation networks, the two airport networks (#52: US Airport network (Batgelj & Mrvar, 2006) and #63: EU Airport network (Cadrillo et al., 2013) got classified as core-heavy; whereas, the road and rail networks (#32: Euro Road network (Subelj & Bajee, 2011) and #70: London Train network (Domenico et al., 2014)) got classified as peripheral-heavy.

Table 1. Domains of the Real-World Networks: Core-heavy vs. Peripheral-heavy Networks

Domain	**Core-heavy Networks**	**Count**	**Peripheral-heavy Networks**	**Count**
Biological	**16, 51, 56, 59, 60, 71, 72, 73**	8	**3, 10, 45, 61, 78, 79, 80**	7
Co-appearance	**13, 53, 66, 68**	4	**55, 69**	2
Employment	**7, 64**	2	**1, 2, 18, 35, 40, 43**	6
Friendship	**8, 9, 17, 22, 23, 26, 36, 38, 48**	9	**24, 62**	2
Literature	**6, 12, 21, 37, 39, 47, 50, 54, 57, 65, 67**	11	**33, 34, 76**	3
Social	**15, 20, 25, 28, 30, 42, 44, 46, 49, 74, 75**	11	**4, 14, 19, 29, 41, 58, 77**	7
Transportation	**52, 63**	2	**32, 70**	2
Miscellaneous	**27**	1	**5, 11, 31**	3

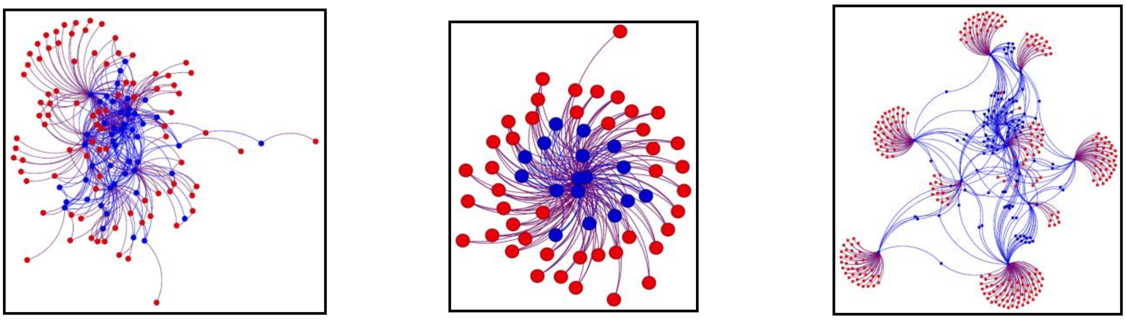
4.4 Core-Peripheral Visualizations of some Benchmark Networks

This sub section presents the visualization of the core-peripheral structure for some benchmark networks commonly studied and analyzed in the Network Science literature. Figure 11 presents the Yifan-Hu proportional layout (Gephi, 2011)-based visualization (the core nodes are colored in blue and the peripheral nodes are colored in red) of the PC_NBNC-based core-peripheral structure determined for six of these benchmark networks: the Band Jazz network (Geiser & Danon, 2003): (#25 in Figure 6), the Karate network (Zachary, 1977): (#20 in Figure 6) and the US Airports '97 network (Batagelj & Mrvar, 2006): (#52 in Figure 6) for core-heavy networks, and the Anna Karnenina network (Knuth, 1993): (#70 in Figure 5), the CS Department Aarhus network (Magnani et al., 2013): (#2 in Figure 5) and the London train transportation network (Domenico et al., 2014): (#70 in Figure 6) for peripheral-heavy networks.



#25: Band Jazz Network #20: Karate Network #52: US Airport Network

Core-heavy Networks



#69: Anna Karnenina Net. #2: CS Dept. Aarhus Net. #70: London Transportation Network

Peripheral-heavy Networks

Figure 11. Visualization of the PC_NBNC-based Core-Peripheral Structures for some Benchmark Real-World Networks [blue colored nodes: core nodes; red colored nodes: peripheral nodes]

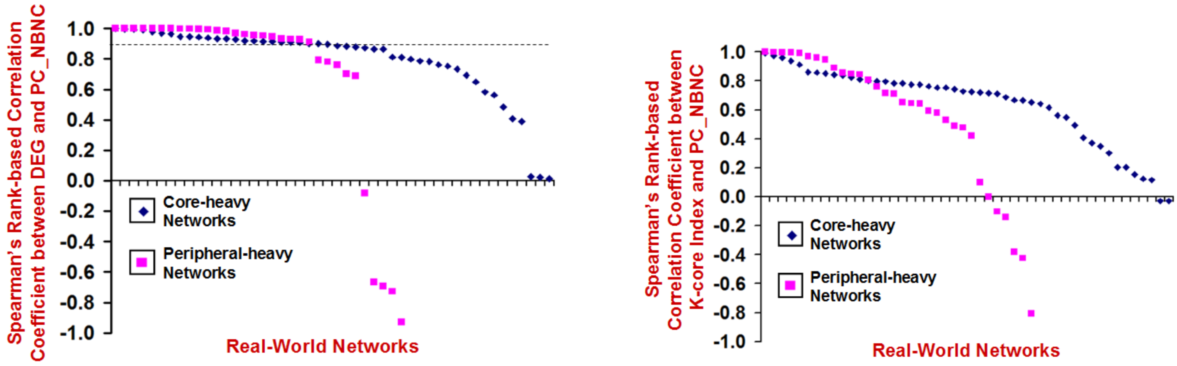
All the three core-heavy networks visualized in Figure 11 are classified so due to them having a larger fraction (more than 50%) of links as core-core links: Band Jazz Network (61.31% core-core links, with 55.05% core nodes); Karate network (53.85% core-core links, with exactly 50% core nodes) and the US Airports network (69.29% core-core links, with only 40.36% nodes as core nodes). All these three networks justify their classification as core-heavy owing to the dense distribution of the core-core links, irrespective of the fraction of core nodes vis-a-vis peripheral nodes. All the three peripheral-heavy networks visualized in Figure 11 are classified so due to their relatively larger fraction of peripheral nodes compared to the core nodes: Anna Karnenina network (64.49% peripheral nodes, with 47.46% of its links being core-peripheral links); the CS Dept Aarhus network (72.13% peripheral nodes, with 69.41% of its links being core-peripheral links) and the London Train Transportation network (70% peripheral nodes, with 52.66% of its links core-peripheral links). Both the CS Dept Aarhus network and the London Train Transportation network do not include any peripheral-peripheral links; also, only 3% of the links in the Anna Karnenina network are peripheral-peripheral, all of which justify their classification as peripheral-heavy networks.

4.5 Correlation Analysis

Figure 12-(i) presents the distribution of the sorted Spearman's correlation coefficient values between DEG and PC_NBNC for both the core-heavy and peripheral-heavy real-world networks. There are 10 out of the 32 peripheral-heavy networks (that is, about 30% of the peripheral-heavy networks) with Spearman's correlation coefficient less than 0.90 (ranging from -0.93 to 0.79) as well as there are 25 out of the 48 core-heavy networks (that is, about 50% of the core-heavy networks) with Spearman's correlation coefficient less than 0.90 (ranging from 0.01 to 0.89).

Figure 12-(ii) presents the Spearman's rank-based correlation between the *k*-core index of the nodes and the PC_NBNC values of the nodes determined for the core-heavy and peripheral-heavy real-world networks of Section 3. The median of the Spearman's rank-based correlation coefficients are 0.72 and 0.68 respectively for the 48 core-heavy networks and for the 32 peripheral-heavy networks. The *k*-core measure has the weakness of overlooking nodes with moderate degree (like node 2 in the toy example graph of Figure 1) but whose neighborhood would otherwise be disconnected without the node. With respect to both DEG and *k*-core index, the Spearman's rank-based correlation coefficients of these two metrics vs. PC_NBNC for the peripheral-heavy

networks are either too high or too low compared to the moderate values observed for the core-heavy networks.

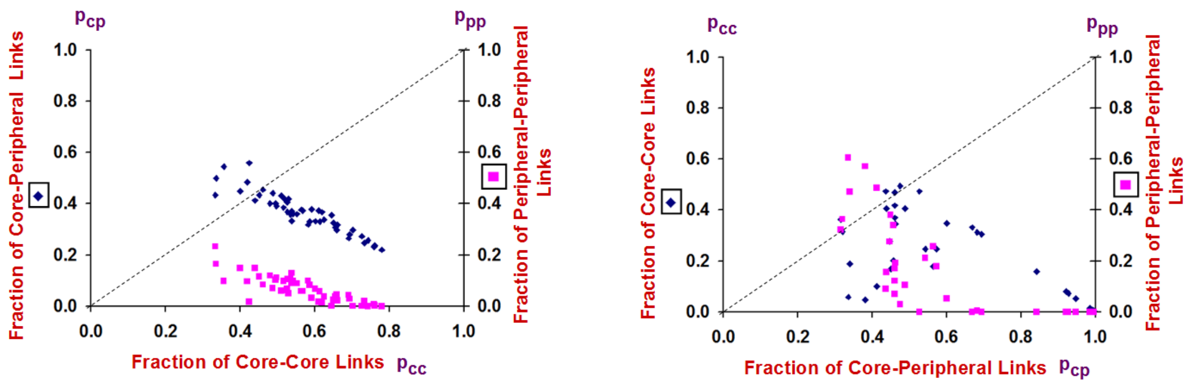


(i) DEG vs. PC_NBNC (ii) *k*-Core Index vs. PC_NBNC

Figure 12. Spearman's Rank-based Correlation Analysis

4.6 Analysis of the Fractions of Links

Figures 13-i and 13-ii plot the actual fractions of core-core links (corresponding to p_{cc}), core-peripheral links (corresponding to p_{cp}) and peripheral-peripheral links (corresponding to p_{pp}) observed respectively for the core-heavy networks and peripheral-heavy networks. In Figure 13-i (plotting p_{cc} vs. p_{cp} and p_{pp}), the above inequality to be primarily satisfied by 42 of the 48 core-heavy networks (only 6 data points lie above the diagonal line in Figure 13-i, corresponding to core-heavy networks for which the fraction of core-core links is not the largest of the three fractions of links) and by only 4 of the 32 peripheral-heavy networks. On the other hand, in Figure 13-ii (plotting p_{cp} vs. p_{cc} and p_{pp}), p_{cp} is primarily the largest of the three fractions of links for 23 of the 32 peripheral-heavy networks (only 9 data points lie above the diagonal line in Figure 13-ii, corresponding to peripheral-heavy networks for which the fraction of core-peripheral links is not the largest) and p_{cc} is the largest of the three fractions of links for only 4 of the 32 peripheral-heavy networks.



(i) Core-heavy Networks (ii) Peripheral-heavy Networks

Figure 13. Validation of the Inequality $p_{cc} > p_{cp} > p_{pp}$ for the Real-World Networks

5. Conclusions and Future Work

The high-level contributions of this paper are two-fold: (1) Principal component analysis (PCA) of the NBNC tuple dataset of the nodes in a complex network and the proposal of a PC_NBNC scalar centrality metric to capture the extent to which nodes play the role of bridge nodes. (2) Classify nodes as core nodes and peripheral nodes based on their PC_NBNC values and accordingly classify a network as core-heavy or peripheral-heavy based on the fractions of the core-core links, core-peripheral links and peripheral-peripheral links as well as the fractions of the core nodes and peripheral nodes. conducted our analysis on a suite of 80 real-world networks and observed 48 of the 80 networks to be core-heavy and the remaining 32 networks to be peripheral-heavy. observed the core-heavy classification of real-world networks to be predominantly driven by the fraction of core-core links (with at least 50% of the links being core-core links in 35 of the 48 core-heavy real-world networks), whereas the peripheral-heavy classification of real-world networks to be predominantly driven by the fraction of peripheral nodes (all the 32 peripheral-heavy networks have more than 50% of their nodes to be peripheral-heavy). observe networks with lower variation in node degree (with the spectral radius ratio for node degree < 2) to be more likely core-heavy (with a probability of 2/3). also observe networks with degree-based assortativity index greater than 0.20 more likely to be core-heavy and those with degree-based assortativity index lower than -0.50 more likely to be peripheral-heavy. As part of future work, plan to investigate the use of clustering algorithms on the PC_NBNC values of the core nodes and peripheral nodes to extract a multi-core/multi-peripheral architecture for complex networks. Also, there is a plan to run a heuristic (Inza et al., 2023) for determining dominating set (Cormen et al., 2009) based on the PC_NBNC values of the nodes to extract a minimal grouping of the nodes, each of which would be part of overlapping clusters comprising of their neighbor nodes. Such overlapping clusters could be further pruned by assessing the contributions of the constituent nodes in the overall modularity (Newman, 2006) of the clusters.

Acknowledgments

Not applicable

Authors' contributions

Not applicable

Funding

This work is partly supported through a sub contract received from University of Virginia titled: Global Pervasive Computational Epidemiology, with the National Science Foundation as the primary funding agency. The views and conclusions contained in this paper are those of the authors and do not represent the official policies, either expressed or implied, of the funding agency.

Competing interests

Not applicable

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Canadian Center of Science and Education.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

Barbera, P., Wang, N., Bonneau, R., Jost, J. T., Nagler, J., Tucker, J., & Gonzalez-Bailon, S. (2015). The Critical Periphery in the Growth of Social Protests. *PLoS ONE*, 0143611, 1-15. https://doi.org/10.1371/journal.pone.0143611

Batagelj, V., & Mrvar, A. (2006). *Pajek Datasets*. Retrieved from http://vlado.fmf.uni-lj.si/pub/networks/data/

Borgatti, S. P., & Everett, M. G. (2000). Models of Core/Periphery Structures. *Social Networks, 21*(4), 375-395. https://doi.org/10.1016/S0378-8733(99)00019-2

Cadrillo, A., Gomez-Gardenes, J., Zanin, M., Romance, M., Papo, D., Pozo, F., & Boccaletti, S. (2013).

Emergence of Network Features from Multiplexity. *Scientific Reports, 3*(1344), 1-6. https://doi.org/10.1038/srep01344

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2022). *Introduction to Algorithms* (4th ed.) MIT Press.

Cucuringu, M., Rombach, P., Lee, S. H., & Porter, M. A. (2016). Detection of Core-Periphery Structure in Networks using Spectral Methods and Geodesic Paths. *European Journal of Applied Mathematics*, *27*, 846-887, 2016. https://doi.org/10.1017/S095679251600022X

De Domenico, M., Sole-Ribalta, A., Gomez, S., & Areans, A. (2014). Navigability of Interconnected Networks under Random Failures. *Proceedings of the National Academy of Sciences*, *111*, 8351-8356. https://doi.org/10.1073/pnas.1318469111

Della Rossa, F., Dercole, F., & Piccardi, C. (2013). Profiling Core-Periphery Network Structure by Random Walkers. *Scientific Reports*, *3*(1467), 1-8. https://doi.org/10.1038/srep01467

Gallagher, R. J., Young, J. G., & Welles, B. F. (2021). A Clarified Typology of Core-Periphery Structure in Networks. *Science Advances*, *7*(12), eabc9800, 1-11. https://doi.org/10.1126/sciadv.abc9800

Geiser, P., & Danon, L. (2003). Community Structure in Jazz. *Advances in Complex Systems*, *6*(4), 563-573. https://doi.org/10.1142/S0219525903001067

Gephi, (2011). Retrieved from https://gephi.org/tutorials/gephi-tutorial-layouts.pdf

Hashler, M., Peienbrock, M., & Doran, D. (2019). dbscan: Fast Density-based Clustering with R. *Journal of Statistical Software*, *91*(1), 1-30. https://doi.org/10.18637/jss.v091.i01

Inza, E. P., Vakhania, N., Sigatreta, J. M., & Mira, F. A. H. (2023). Exact and Heuristic Algorithms for the Domination Problem. *European Journal of Operational Research*, *313*(2), 1-30. https://doi.org/10.1016/j.ejor.2023.08.033

Jolliffe, I. T. (2002). *Principal Component Analysis* (1st Ed.), Springer Series in Statistics.

Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010). Identification of Influential Spreaders in Complex Networks. *Nature Physics*, *6*(11), 888-893. https://doi.org/10.1038/nphys1746

Knuth, D. E. (1993). *The Stanford GraphBase*: *A Platform for Combinatorial Computing* (1st Ed.) Addison-Wesley.

Kojaku, S., & Masuda, N. (2017). Finding Multiple Core-Periphery Pairs in Networks. *Physical Review E*, *96*, 052313. https://doi.org/10.1103/PhysRevE.96.052313

Lloyd, S. (1982). Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129-137. https://doi.org/10.1109/TIT.1982.1056489

Ma, C., Xiang, B. B., Chen, M. S., Small, M., & Zhang, H. F. (2018). Detection of Core-Periphery Structure in Networks based on 3-Tuple Motifs. *Chaos*, *28*, 053121. https://doi.org/10.1063/1.5023719

Magnani, M., Micenkova, B., & Rossi, L. (2013). *Combinatorial Analysis of Multiple Networks.* arXiv:1303.4986 [cs.SI].

Maia de Abreu, N. M. (2007). Old and New Results on Algebraic Connectivity of Graphs. *Linear Algebra and its Applications*, *423*(1), 53-73. https://doi.org/10.1016/j.laa.2006.08.017

Meghanathan, N. (2012). Neighborhood-based Bridge Node Centrality Tuple for Complex Network Analysis. *Applied Network Science*, *6*(47), 1-36. https://doi.org/10.1007/s41109-021-00388-1

Meghanathan, N. (2014). *Spectral Radius as a Measure of Variation in Node Degree for Complex Network Graphs*. Paper presented at the 3rd International Conference on Digital Contents and Applications, Hainan, China. https://doi.org/10.1109/UNESST.2014.8

Meghanathan, N. (2016). Assortativity Analysis of Real-World Network Graphs based on Centrality Metrics. *Computer and Information Science*, *9*(3), 7-25. https://doi.org/10.5539/cis.v9n3p7

Meghanathan, N. (2017). Randomness Index for Complex Network Analysis. *Social Network Analysis and Mining*, *7*(25), 1-15. https://doi.org/10.1007/s13278-017-0444-3

Meghanathan, N. (2024). Assortativity Analysis of Complex Real-World Networks using the Principal Components of the Centrality Metrics. *International Journal of Data Science*, *9*(1), 79-97.

https://doi.org/10.1504/IJDS.2024.135945

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

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

Puck Rombach, M., Porter, M. A., Fowler, J. H., & Mucha, P. J. (2014). Core-Periphery Structure in Networks. *SIAM Journal on Applied Mathematics*, *74*(1), 167-190. https://doi.org/10.1137/120881683

Strang, G. (2023). *Linear Algebra and its Applications*. (6th Ed.) Wellesley-Cambridge Press.

Subelj L., & Bajec, M. (2011). Robust Network Community Detection using Balanced Propagation. *The European Physical Journal B*, *81*(3), 353-362. https://doi.org/10.1140/epjb/e2011-10979-2

Yanchenko, E., & Sengupta, S. (2023). Core-Periphery Structure in Networks: A Statistical Exposition. *Statistics Surveys*, *17*, 42-74. https://doi.org/10.1214/23-SS141

Zachary, W. W. (1977). An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research*, *33*(4), 452-473. https://doi.org/10.1086/jar.33.4.3629752

Zhang, X., Martin, T., & Newman, M. (2015). Identification of Core-Periphery Structure in Networks. *Physical Review E*, *91*(3), 032803. https://doi.org/10.1103/PhysRevE.91.032803