Graph clustering using correlation between rankings of centrality measures

Alice Nanyanzi
Mentors:
Prof Renaud Lambiotte and Karel Devriendt
Mfano Africa-Oxford Virtual Mentorship Programme

Centrality is one of the vital concepts in network theory. The centrality of a node is a measure of its 'importance' in a network. There are different centrality measures used to quantify different importances of nodes and while some correlate and thus show a similar notion of centrality, in other cases they do not correlate. The application of these measures is guided by the nature of the network and the empirical phenomenon of interest. In this work, we investigate how structural similarity between networks can be obtained using the correlations between rankings of selected centrality measures. The similarity metric that we used is the distance between centrality correlation matrices of some real-world networks from which clustering of the networks follows.

INTRODUCTION

Centrality of nodes plays a key role in the study of networks as nodes are considered important based on how quickly they reach other nodes in the network, how many neighbours they have, their role as a bridge among pairs of nodes, among others. The different 'importances' transform to the real-world in different ways for instance, the influence of a person on a social network, a super-spreader of a disease in a population, among others. The concept of centrality is ambiguous hence the numerous centrality measures which are clearly illustrated as a periodic table in [1]. In this work, however, we will focus on some of the most popular metrics namely degree, betweenness, closeness, eigenvector, subgraph, communicability betweenness, harmonic, load, effective resistance closeness and betweenness centralities. These metrics are categorised based on shortest paths, walks, random walks and non-back tracking walks as reviewed in [2].

Before we continue the discussion of centrality measures, let us briefly present some preliminary notions on graphs.

Preliminaries

A graph is a pair $G = (V, E)$, where V is a set of vertices or nodes, and E is a set of edges between the vertices, $E \subseteq \{(u, v)|u, v \in V\}$ [3]. A graph may be undirected, that is its edges have no orientation or it may be directed, that is its edges have direction. In this work, we consider undirected graphs unless otherwise stated. A graph $G' = (V', E')$ is a subgraph of G if and only if $V' \subseteq V$ and $E' \subseteq E$. The degree of a node, k_v, is the number of edges connected to it. The average degree is the average of the degrees of nodes in a graph.

A path is a sequence of edges which join a sequence of nodes in which all the nodes and edges are distinct [4]. The length of a path between a pair of nodes is known as its path length. Shortest path distance is the least distance between a pair of nodes while the diameter is the greatest shortest distance between any pair of vertices. The average path length is the average number of steps along the shortest paths for all possible pairs of network nodes. Clustering coefficient is a measure of the degree to which nodes tend to cluster together [5].

Graphs can be represented using different matrices. The most common and insightful ones are the adjacency and Laplacian matrices. They capture the connections within a graph and their properties give useful information about structure of the graphs they represent. For a graph with n nodes, its adjacency matrix, $\mathbf{A}$, is an $n \times n$ matrix whose entries are given by [3]

$$\mathbf{A}_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The Laplacian matrix, $\mathbf{L}$, is defined as [3]

$$\mathbf{L}_{ij} = \begin{cases} k_i & \text{if } i = j, \\ -1 & \text{if } (i,j) \in E \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The Laplacian matrix is real, symmetric and positive semi-definite [6]. These and more properties make the Laplacian matrix desirable as interesting structural properties of the associated graphs are revealed using matrix properties.

CENTRALITY METRICS BASED ON SHORTEST PATH DISTANCE

The shortest distance between a pair of nodes in an undirected unweighted graph is the minimum number of edges in a path connecting the pair. For unconnected graphs, the shortest distance between a pair of nodes in different components is infinity. Centrality measures based on shortest path distance include but not limited to closeness, harmonic, betweenness, and load centralities.

Closeness centrality

As suggested by its name, the closeness centrality of a node quantifies how close a node is, to the other nodes in a graph. Closeness is defined as the reciprocal of farness, as

$$C_c(i) = \frac{1}{\sum_{i \neq j} d_{i,j}}, \tag{3}$$

where $d_{i,j}$ is the shortest path distance between the node pair i and j. The closeness centrality is normalized by $(n-1)$ where n is the number of nodes in the graph [7]. It is worth noting that for unconnected graphs, the closeness centrality of all nodes is zero. To circumvent this, a reformulation of (3) was presented in [8].It is feferred to as harmonic centrality.

Harmonic centrality

Harmonic centrality is a variant of closeness centrality that works well with unconnected graphs. Harmonic centrality of a node i is the sum of the reciprocal of the shortest path distances from all other nodes to i and it is given by [8]

$$H_c(i) = \sum_{i \neq j} \frac{1}{d_{i,j}}. \tag{4}$$

Betweenness centrality

Betweenness is a centrality metric that quantifies the brokering power of node. In other words, it measures the power of a node to intercept communications between pairs of nodes in a graph, that is, the fraction of shortest paths going through a particular node. It is defined as

$$B_c(i) = \sum_{j,k \in V} \frac{\sigma(j,k|i)}{\sigma(j,k)}, \tag{5}$$

where $\sigma(j,k|i)$ is the number of shortest paths between node j and k that pass through node i and $\sigma(j,k)$ is the total number of shortest paths between the node pair [9]. The betweenness values are normalized by $2/(n-1)(n-2)$ where n is the number of nodes in the graph.

The concept of betweenness centrality is also extended to edge betweenness which quantifies the role of an edge as a bridge in the network.

Load centrality

Load centrality is a betweenness-like measure based on hypothetical flow process, that is, suppose each vertex sends a unit quantity of a commodity to every other vertex to which it is connected, with routing based on a priority system. For instance, consider a flow with node j as the source and node k as the destination then the quantity x flowing into node i is dispersed equally to its neigbours that have a minimum geodesic distance to the destination vertex k [10–12]. The load centrality node i is the total fraction of commodity flowing through it for all pairs of source and target nodes. Mathematically, load centrality of a node i is given by

$$L_c(i) = \sum_{j,k \in V, i \neq j \neq k} \theta_{j,k}(i), \tag{6}$$

where $\theta_{j,k}(i)$ is the overall commodity flowing into node i and $\theta_{j,k} = 1$ is the quantity of commodity sent from source node j to target k. The load centrality values are normalized by $2/(n-1)(n-2)$ where n is the number of nodes in the graph.

CENTRALITY MEASURES BASED ON WALKS

A walk in a network is a series of edges (not necessarily distinct)

$$(u_1, v_1), (u_2, v_2), \ldots, (u_k, v_k),$$

for which $v_i = u_{i+1}$ $(i = 1, 2, \ldots, k-1)$ [5]. A walk of length k is referred to as a k-walk. We can compute the number of k-walks between any pair of nodes using the entries of $\mathbf{A}^k$, more precisely, $(\mathbf{A}^k)_{ij}$ equals the number of k-walks between i and j.

Degree centrality

Degree is a measure of centrality based on the number of direct neighbours of a node. In terms of walks, degree is the number of walks of length one with node i as an end-point. It is given as

$$D_c(i) = \sum_{j \in N} \mathbf{A}_{ij}, \tag{7}$$

where N denotes the set of neighboring nodes [13]. It is considered the simplest measure of centrality and quite informative though for some networks where a large number of nodes have the same degree then the measure may not be very insightful.

Eigenvector centrality

The eigenvector centrality is a measure of the influence of a node in a network. It takes into account the quality of neighboring nodes by assigning relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes [14, 15]. Suppose that we have an initial centrality score for each node i as $x_i(0)$. After a propagation of walks of length l then the centrality vector is given by

$$\mathbf{x}(l) = \mathbf{A}^l \mathbf{x}(0) = \sum_i c_i \lambda_i^l \mathbf{v}_i, \tag{8}$$

where $\mathbf{v}_i$ is the eigenvector associated with the i-th eigenvalue, λ_i, of $\mathbf{A}$ [2]. As $l \longrightarrow \infty$, the centrality converges to leading eigenvector which is associated with the largest eigenvalue of $\mathbf{A}$ [13]. Since $\mathbf{A}$ is real square matrix with non-negative entries and if G is connected, by Perron-Frobenius theorem, this eigenvector is positive. Thus the vector of the eigenvector centralities of the nodes is proportional to the leading eigenvector. The eigenvector centrality of vertex i can be defined as:

$$E_c(i) \propto \sum_j \mathbf{A}_{i,j} E_c(j). \tag{9}$$

In matrix form, we have

$$\mathbf{e} = \mathbf{A}\mathbf{e}, \tag{10}$$

where $\mathbf{e}$ is a vector of eigenvector centralities for all nodes, that is, $\mathbf{e}_i = E_c(i)$.

Communicability betweenness

Communicability betweenness is a variant of betweenness centrality that measures the controllability of a node in consideration of walks of all lengths. More precisely, it is the fraction of all walks through the node and is defined as [16]

$$CB_c(i) = \frac{1}{C} \sum_j \sum_k \frac{w(j,k|i)}{w(j,k)}, \tag{11}$$

where $w(j,k|i)$ is the number of walks between pair (j,k) that pass through node i, $w(j,k)$ is the total number of walks between the node pair and C is a normalization factor.

Subgraph centrality

Subgraph centrality characterises the importance of a node by its participation in all subgraphs in the graph. The subgraph centrality of a node i is the sum of weighted closed walks of all lengths starting and ending at node i. The weights decrease with path length. It is given as

$$S_c(i) = \sum_{k=0}^{\infty} \frac{u_k(i,i)}{k!} = \sum_{k=0}^{\infty} \frac{(\mathbf{A}^k)_{ii}}{k!} = (e^{\mathbf{A}})_{ii}, \tag{12}$$

where $u_k(i,i)$ is the number of closed k-walks that vertex i participates in, in the network [17].

CENTRALITY MEASURES BASED ON EFFECTIVE RESISTANCE

Like the shortest path distance, the effective resistance is a distance measure based on paths between nodes in a graph, however, it takes into account all paths and their lengths not the shortest paths only. The definition of the effective resistance of a graph is anchored on the pseudo-inverse of the Laplacian matrix which exists given that the Laplacian matrix has one zero eigenvalue with its associated eigenvector $\mathbf{v} = (1, \dots, 1)^T$. Thus the effective resistance, $\Omega_{i,j}$ between a pair of nodes is given by

$$\Omega_{i,j} = (e_i - e_j)^T \mathbf{L}^\dagger (e_i - e_j), \tag{13}$$

where the unit vectors have entries $(e_i)_k = 1$ if $k = i$ and zero otherwise [18, 19].

Since the effective resistance is a distance measure, the concept was extended to centrality measure such as in closeness and betweenness centralities where the effective resistance is used rather than the shortest path distance resulting in the effective resistance closeness and betweenness centralities. These centrality measures, derived from the flow of current through an electrical network, relax the assumption that information flows only through shortest paths.

Effective resistance closeness centrality

Effective resistance closeness centrality also known as information centrality or current-flow closeness centrality is a variant of closeness centrality based on effective resistance between pairs of nodes. It is given by

$$EC_c(i) = \frac{n-1}{\sum_{i \neq j} \Omega_{i,j}}, \tag{14}$$

where n is the number of nodes in a graph [20].

Effective resistance betweenness centrality

The effective resistance betweenness centrality, also known as current-flow betweenness, of a vertex i is the amount of current that flows through i averaged over all pairs of source and target nodes [20–22]. It is given by

$$EB_c(i) = \frac{\sum_{s<t} F_i^{(s,t)}}{(1/2)n(n-1)}, \tag{15}$$

where

$$F_i^{(s,t)} = 1/4 \sum_{i \neq j} A_{i,j} |\Omega_{i,s} - \Omega_{j,s} + \Omega_{j,t} - \Omega_{i,t}|, \tag{16}$$

and s is the source node, t is the target node.

GRAPH CLUSTERING USING CENTRALITY CORRELATION

Graph clustering is a key concept that has application in various fields. Graph clustering can take on two forms namely within graph clustering and between graph clustering. The former involves grouping of nodes within a single graph while the latter entails grouping of graphs based on some kind of similarity. In this work, we consider the latter in which the similarity measure is based on the correlations of the rankings of centrality measures.

Correlation between centrality metrics

Given the key role of centrality in network analysis, it is equally paramount to ascertain the correlation of the rankings of different centrality measures for various reasons: firstly, to identify less computationally expensive metrics in case redundancy between measures is detected [23]. Secondly, correlation aids in justifying the novelty of a new centrality metric, that is, a weak correlation of a new metric with existing ones is an indicator that the new measure captures importance on a different structural level to the existing metrics which proves its novelty [21].

The other reason which is the pivot of this work is the use of correlation between metrics to ascertain the level of similarity between graphs [24]. For a pair of graphs, we compute the distance between their corresponding centrality correlation matrices and the smaller the distance, the more structurally similar the two graphs are.

We are interested in the rankings of indices and as such we use rank-based correlation measures. We use the Kendall's tau-b as it is independent of the distribution of the data, reflects the strength of dependence between variables being tested and also accounts for ties if any. The Kendall's tau-b, τ is defined by

$$\tau = \frac{P-Q}{\sqrt{(P+Q+T)\times(P+Q+U)}}, \tag{17}$$

where P is the number of concordant pairs, Q the number of discordant pairs, T the number of ties only in x, and U the number of ties only in y [25]. If a tie occurs for the same pair in both x and y, it is not added to either T or U.

Figure 1 shows the Zachary karate club network and its correlation matrix of the 10 centrality measures.

Graphs are classified as similar (i.e will belong to the same cluster) if their rankings of the 10 metrics is similar.

The clustering process involves the following steps:

1. Given a set of m graphs $\{G_1, G_2, \ldots, G_m\}$, compute the rankings of the 10 centrality metrics for each graph.

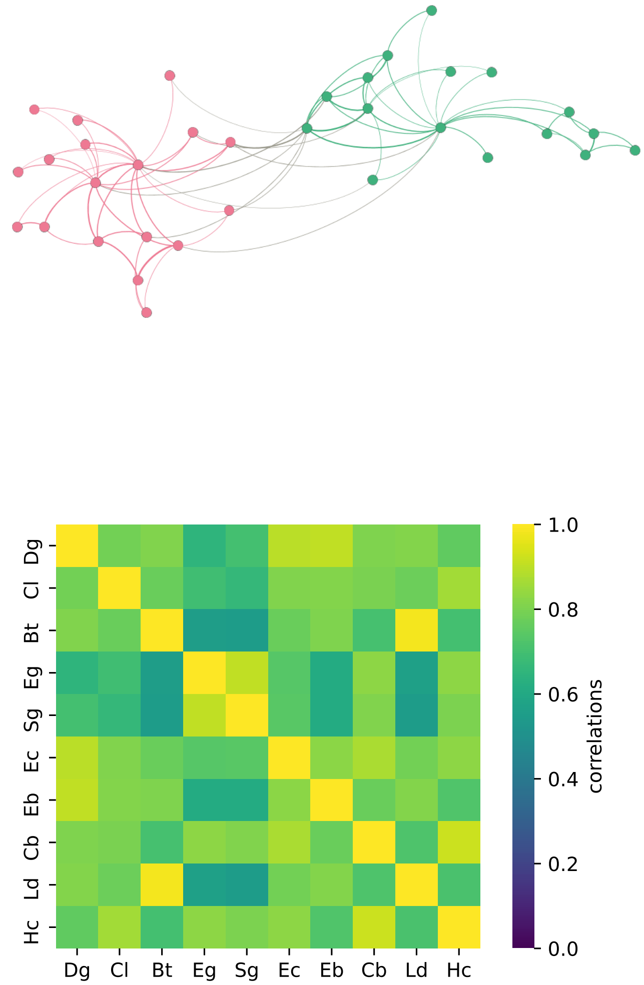

FIG. 1: The Zachary karate club network and its correlation matrix (as a heatmap) of the rankings of the 10 centrality measures.

2. Obtain the correlation matrix between the rankings for each graph. We used Kendall's correlation for this work.

3. Obtain the distance among the matrices representing different graphs. This forms a $m \times m$ distance matrix.

4. Perform clustering using the distance matrix.

For the experiments, we worked with 18 real-world networks categories as biological, social, ecological, brain, software, informational, technological and infrastructural networks. Specifically, the networks are Neurons C. elegans, Canton Creek, El Verde, Scotch Broom, Ythan_1,

Name	Nodes	Edges	Ave.deg	Diam	Ave.pl	co-eff
ElVerde	156	1441	18.474	5	2.300	0.214
SWCitations	233	994	8.532	4	2.371	0.556
LittleRock	181	2328	25.724	5	2.225	0.342
Stony	112	832	14.857	4	2.336	0.068
Ythan1	134	597	8.910	4	2.403	0.228
Celegans	280	1973	14.093	6	2.630	0.280
Canton	108	708	13.111	5	2.350	0.053
USAir97	332	2126	12.807	6	2.738	0.625
GD	249	635	5.100	11	4.150	0.239
PINHpylori	710	1396	3.932	9	4.152	0.015
PINMalaria	229	604	5.275	8	3.375	0.175
Digital	150	198	2.640	12	4.847	0.051
Yeast	662	1062	3.208	15	5.200	0.049
Drugs	616	2012	6.532	13	5.284	0.550
Electronics2	252	399	3.167	13	5.806	0.056
Ecoli	328	456	2.780	13	4.834	0.110
Zachary	34	78	4.588	5	2.408	0.571
SBroom	154	370	4.805	6	3.394	0.141

TABLE I: Summary statistics for the 18 datasets.

Stony Stream, Small World, PPI Malaria, PPI H. pylori, Trans_E.coli, Trans_yeast, Drugs, Electronic2, USAir97, GD, Digital and Zachary (properties in Table I). Figure 1 is the Zachary karate club graph and its correlation heat map for the 10 centrality measures and Figure 2 shows the correlation matrices for all the 18 networks.

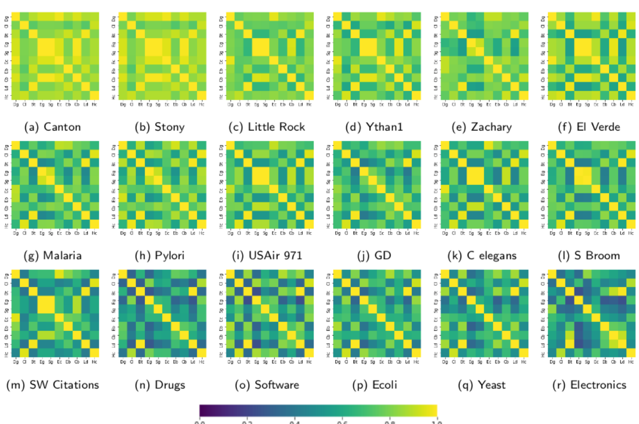

FIG. 2: Correlations of the rankings of the 10 centralities for the 18 real-world networks.

From Figure 2, we observe three categories of graphs based on the correlations of the subgraph and eigenvector centrality with the rest of the centralities. For networks in category A, the rankings of the correlations of the subgraph and eigenvalue centralities with the other centralities is similar. For this reason, one would choose to consider eigenvector centrality as it is computationally

Category A	Category B	Category C
Canton	Zachary	GD
Stony	PINMalaria	Digital
Little Rock	PINHpylori	Drugs
Ythan1		Ecoli
El Verde		Electronics
Citations		Yeast
C elegans		
USAir97		
Scotch Broom		

TABLE II: Networks belonging to each of the 3 categories

less expense than subgraph centrality. Networks in category B only show a slight difference while for category C the subgraph and eigenvector centralities correlate differently with the rest of the centralities. Figure 3 is the visualisation of the 3 categories based on this interpretation and Table II shows which networks belong to each category.

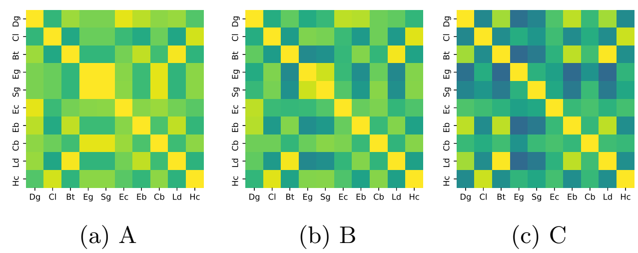

FIG. 3: The 3 categories based on subgraph and eigenvector centralities correlation with the other 8 centrality metrics.

With the correlation matrices handy, we used the Euclidean norm (L^2 norm) to obtain the distance between matrices. Given a pair of matrices $M1$ and $M2$, the distance is given as [26]

$$D_{i,j} = \sqrt{(M1_{i,j} - M2_{i,j})^2}. \tag{18}$$

Figure 4 depicts the distance matrix obtained fo the 10 classical centrality measures. With the distance matrix, we obtain clustering using hierarchical agglomerative clustering algorithm which builds nested clusters by merging them successively to form bigger clusters using a bottom up approach. Each graph in its own cluster and then successfully merges them using ward linkage method that minimizes the variance of the clusters being merged [27]. The resulting clustering is then visualised as a dendrogram in Figure 5.

Taking a threshold distance of 4, we observe two clusters which we refer to as cluster A in green and cluster B in red as depicted in Figure 5. This clustering agrees with the statistical properties of these networks, for example, networks in cluster C namely electronic2, yeast,

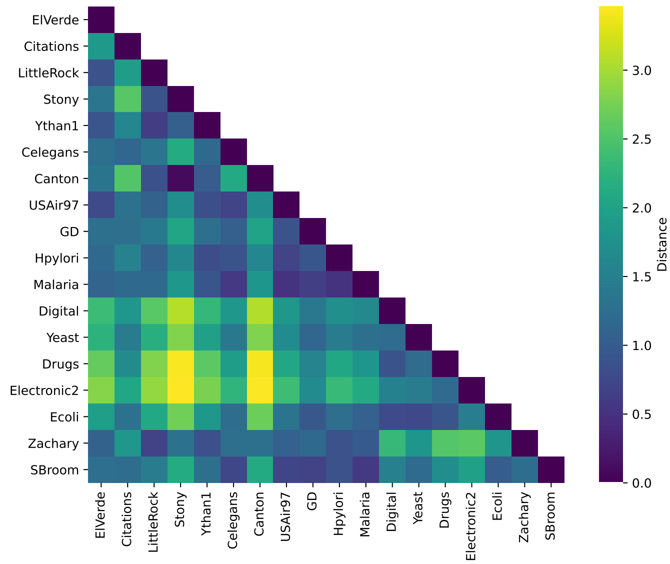

FIG. 4: The distance matrix for real-world networks.

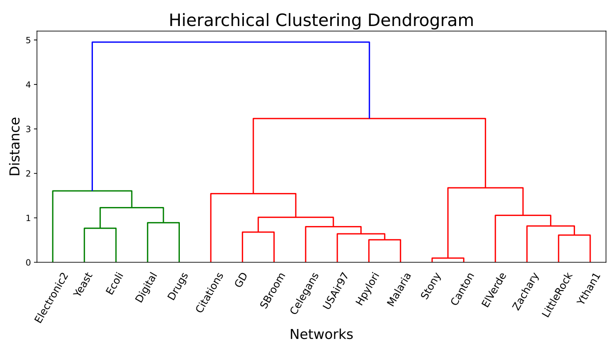

FIG. 5: Dendrogram depicting the clustering of networks

drugs, digital and ecoli have diameter and average path length greater than 11 and 4.5 respectively. Since the 10 classical centrality measures used in the experiments are based on paths and walks, then the statistic properties of the networks in cluster C are partially responsible for the similar patterns of correlation rankings and the observed differences from the rest of other networks that belong to clusters A and B.

CENTRALITY BASED ON THE VARIANCE OF DISTRIBUTIONS ON GRAPHS

Classical node centrality measures ascertain the importance of a node based on its role in the structure of the network. This structural importance does not account for external attributes on the nodes, which is referred to as individual importance [24]. Individual importance may be a key contributor to the overall importance of a node and thus the need to be considered as well. The variance-based centrality introduced in this work ascertains the importance of a node based on both structural and individual importance of the node. Firstly, let us discuss the concept of variance of distributions on graphs.

Variance of distributions on Graphs

In [28], authors introduced the concept to measure the variance of probability distributions on the nodes of graphs. This takes into account the distance between nodes which could be the shortest path distance or the effective resistance.

Given a probability distribution p and distance notion d, the variance of the distribution on a graph is given by

$$\mathrm{var}(p) = \frac{1}{2} \sum_{i,j \in V} p(i)p(j)d^2(i,j). \tag{19}$$

Variance-based centrality

We define the centrality of a node as the contribution of that node to the variance of the distribution, p, on the graph. It is given by

$$\mathrm{C}_v(p,i) = \frac{1}{2} \sum_{j \in V, j \neq i} p(i)p(j)d^2(i,j) \tag{20}$$

From (20), it is evident that a node with the least contribution is the most central. Also, when a uniform distribution is considered across the nodes, then the variance-based centrality with the shortest distance and effective resistance as distance becomes very close to the normal closeness centrality and the effective distance closeness centrality respectively. The only difference is that variance-based centrality with shortest distance takes the square of the distance while the classical closeness takes simply the distance. With this, we expect high correlations between these centrality measures.

Table III shows the correlations of the rankings for the two non-classical centrality measures where the distance metric are effective resistance (Rv) and squared values of shortest path distance (Sv) for 18 networks. The distribution on each graph is uniform and it is given by $p_i = 1/n_i$ where n_i is the number of nodes for a graph i. Figures 6 and 7 show the distance matrix and hierarchical classification of networks.

A first observation is that using the variance-based centralities, the distance between correlation matrices is low (atmost 1.0) compared to the one of the classical centralities. Secondly, the citation network is shown to be different from the rest of the networks possibly because of its properties namely low average shortest path length

Networks	Correlation
ElVerde	0.8232
SWCitations	0.4239
LittleRock	0.7154
Stony	0.7654
Ythan1	0.6575
Celegans	0.6207
Canton	0.7595
USAir97	0.7244
GD	0.7184
Hpylori	0.7362
Malaria	0.6702
Digital	0.7869
Yeast	0.6796
Drugs	0.5934
Electronic2	0.5921
Ecoli	0.7082
Zachary	0.7318
SBroom	0.6794

TABLE III: Correlations between variance-based centralities with shortest path distance and effective resistance as distance metrics.

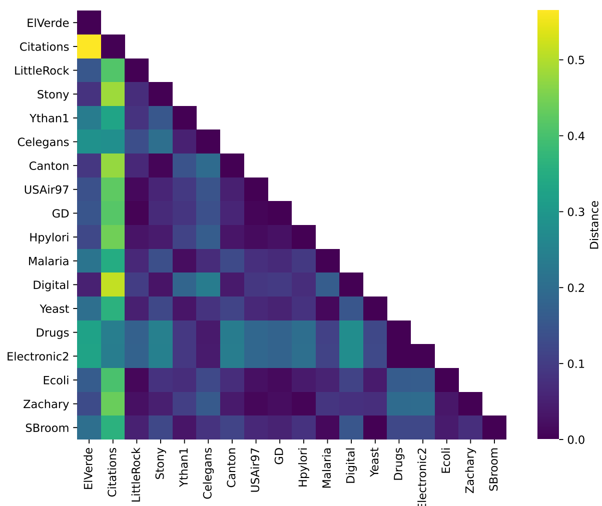

FIG. 6: The variance-based distance matrix for the real-world networks

of 2.371 and a high clustering coefficient of 0.556 in comparison to the rest of the networks. Its small-world coefficients σ and ω are 1.21 and -0.18 respectively which characterises it as a small-world network.

COMBINATION OF CLASSICAL AND VARIANCE-BASED CENTRALITY CLUSTERING

We combine the 10 classical and the 2 variance-based centrality metrics totalling to 12 as shown in Figure 8 for the Zachary karate graph.

We obtained the correlation matrices as shown in Figure 9 followed by the distance matrix and hierarchical clustering in Figures 10 and 11.

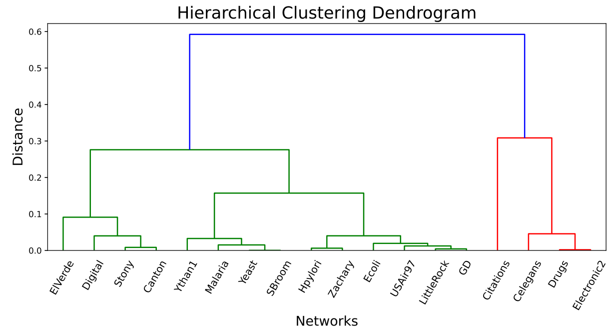

FIG. 7: Dendrogram depicting the clustering of networks

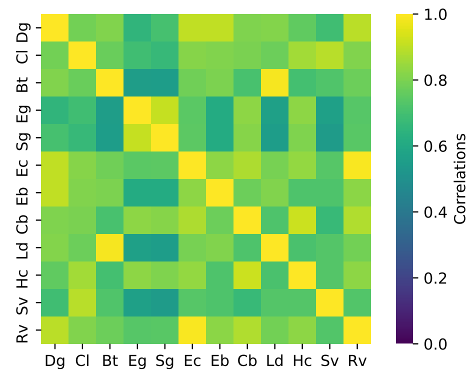

FIG. 8: The combined classical and variance-based correlation matrices for the for the Zachary karate graph

Taking a threshold distance of 4.5, we observe from the dendrogram in Figure 11 that there are two clusters as was the case in the classical centrality clustering shown in Figure 5.

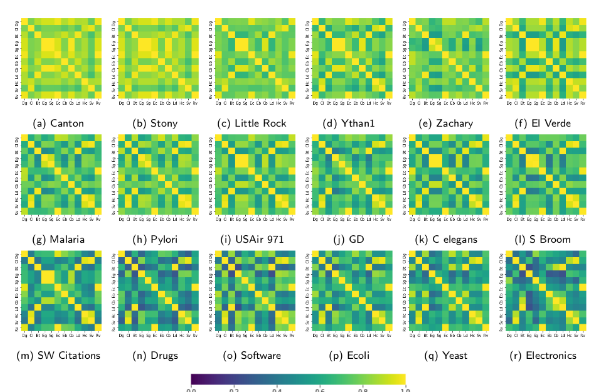

FIG. 9: The combined classical and variance-based correlation matrices for the 18 real-world networks

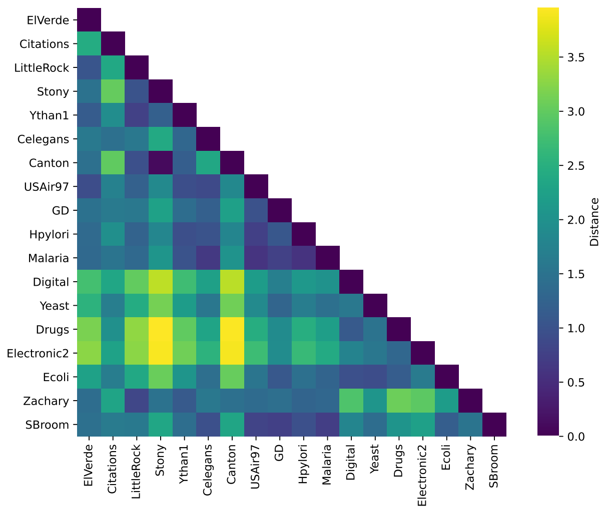

FIG. 10: The combined classical and variance-based distance matrix for real-world networks

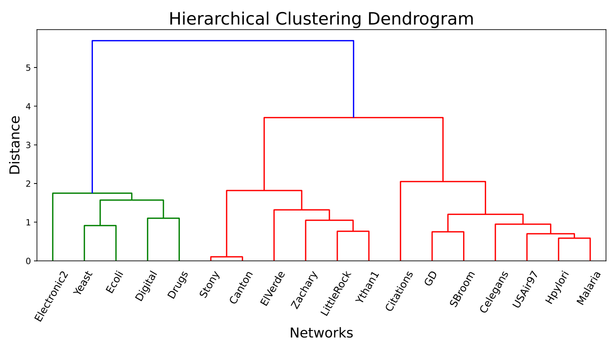

FIG. 11: Dendrogram depicting the clustering of networks

CONCLUSION

We have explored similarity of graphs by taking a group of graphs and perform clustering using node centrality rank correlations. Our experiments were based on 12 real world networks. Firstly, we considered 10 classical measures that are based on the network structure to determine the importance of nodes. Secondly, we considered two variance-based centrality measure with the shortest distance and effective resistance as distance metrics. These measures combine both structure and individual importance of the node to determine the centrality of a node by considering a distribution of a graph. In this study, we considered uniform distributions over the graphs. Results from both experiments give similar hierarchical clustering for the networks which agree with the statistics of the networks.

We hope to extend this work in the following ways: Firstly, ascertaining the validity of clustering by experimenting on different graph models say Erdös-Rényi models versus stochastic block models. Secondly, we will consider non-uniform distributions on the networks and explore how the centrality rankings change hence a change in the clustering of the networks.

[1] D. Schoch, URL http://schochastics. net/sna/periodic. html (2016).
[2] A. Bovet and H. A. Makse, arXiv preprint arXiv:2105.01931 (2021).
[3] M. Newman, *Networks: an introduction* (OUP Oxford, 2010).
[4] E. Estrada, in *Evolutionary Equations with Applications in Natural Sciences* (Springer, 2015) pp. 93–131.
[5] E. Estrada, P. Knight, *et al.*, *A first course in network theory* (Oxford University Press, USA, 2015).
[6] O. R. Oellermann and A. J. Schwenk, (1991).
[7] L. C. Freeman, Social networks **1**, 215 (1978).
[8] P. Boldi and S. Vigna, Internet Mathematics **10**, 222 (2014).
[9] L. C. Freeman, Sociometry , 35 (1977).
[10] K.-I. Goh, B. Kahng, and D. Kim, Physical review letters **87** (2001).
[11] M. E. Newman, Physical review E **64**, 016132 (2001).
[12] L. Maccari, L. Ghiro, A. Guerrieri, A. Montresor, and R. L. Cigno, in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications* (IEEE, 2018) pp. 2582–2590.
[13] M. Newman, *Networks* (Oxford university press, 2018).
[14] L. Katz, Psychometrika **18**, 39 (1953).
[15] P. Bonacich, Journal of mathematical sociology **2**, 113 (1972).
[16] E. Estrada, D. J. Higham, and N. Hatano, Physica A: Statistical Mechanics and its Applications **388**, 764 (2009).
[17] E. Estrada and J. A. Rodriguez-Velazquez, Physical Review E **71**, 056103 (2005).
[18] G. Ranjan and Z.-L. Zhang, Physica A: Statistical Mechanics and its Applications **392**, 3833 (2013).
[19] D. J. Klein and M. Randić, Journal of mathematical chemistry **12**, 81 (1993).
[20] U. Brandes and D. Fleischer, in *Annual symposium on theoretical aspects of computer science* (Springer, 2005) pp. 533–544.
[21] M. E. Newman, Social networks **27**, 39 (2005).
[22] E. Bozzo and M. Franceschet, Social Networks **35**, 460 (2013).
[23] C. Li, Q. Li, P. Van Mieghem, H. E. Stanley, and H. Wang, The European Physical Journal B **88**, 1 (2015).
[24] D. Schoch, *A positional approach for network centrality*, Ph.D. thesis, Universität Konstanz (2015).
[25] W. R. Knight, Journal of the American Statistical Association **61**, 436 (1966).
[26] G. H. Golub and C. F. Van Loan, "Matrix computations. johns hopkins studies in the mathematical sciences," (1996).
[27] A. K. Jain and R. C. Dubes, *Algorithms for clustering data* (Prentice-Hall, Inc., 1988).
[28] K. Devriendt, S. Martin-Gutierrez, and R. Lambiotte, arXiv preprint arXiv:2008.09155 (2020).