



HAL
open science

Structural network analysis: Correlation between centrality measures

Claude Kanyou, Etienne Kouokam, Yves Emvudu

► **To cite this version:**

Claude Kanyou, Etienne Kouokam, Yves Emvudu. Structural network analysis: Correlation between centrality measures. CARI 2022, Oct 2022, Yaoundé, Dschang, Cameroon. hal-03714191

HAL Id: hal-03714191

<https://hal.science/hal-03714191>

Submitted on 12 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

Structural network analysis: Correlation between centrality measures

Claude Kanyou^{*1}, Etienne Kouokam², Yves Emvudu³

^{1,2}Department of Computer Science, Faculty of Science, Université de Yaoundé I, Yaoundé, Cameroon

³Department of Mathematics, Faculty of Science, Université de Yaoundé I, Yaoundé, Cameroon

*E-mail : kanyouclaude@gmail.com

Abstract

Centrality measures are important in network analysis and mining. The correlation structures between the centrality measures are subject of the structural network analysis to improve the understanding of the process in that network. The aim of this paper is to extract useful hidden structural information about the structural correlations between centrality measures study. With Data Science analysis and visualization technics, we propose a structural analysis and visualization of the Pearson correlation between centrality measures. Overall experimentations, the same results are shown in 31 social networks with different topologies. The results show the existence of a strong positive and structural correlation between degree and closeness centrality, degree and betweenness centrality, degree and eigenvector centrality. However, we observed a strong positive and non-structural correlation between eigenvector and betweenness centrality, betweenness and closeness centrality, closeness and eigenvector centrality. Furthermore, we suggest some structural implications of these centrality measures in a network. Finally, we identify influential nodes and their state of evolution that build an effective minimization policy for an infectious disease spread. With the strong positive and structural correlations observed, in large networks, high complexity centrality measure can be approximated by low complexity such as degree centrality.

Keywords

Social Network Analysis and Mining ; Network structures ; Structural analysis ; Centrality Measure ; Structural correlation ; Infectious disease spread.

I INTRODUCTION

Networks are ubiquitous and network science has revolutionized research in the complex system. They are flexible tools for the modeling and studying complex systems [Pellis et al. \(2014\)](#). In the network, various explicit/local or implicit/global links connect the nodes at different structural levels. Thus the centrality measures have interest in the network where the understanding network structure allows to improve a better control of the micro and macro propagation [Keeling and Rohani \(2008\)](#), [Danon et al. \(2011\)](#) and even improve predictions. In public health, the infectious diseases spread through nodes in a contacts network. Some nodes have a high-spreading capacity than others. Thus, the *Social Network Analysis and Mining (SNAM)* is useful for the modelling and the study of the network by typical analysis based on centrality measures. This in order to identifying influencer and structural nodes to build a policy to mitigation the propagation, [Enright and Kao \(2018\)](#), [Pellis et al. \(2014\)](#), [Keeling and Rohani \(2008\)](#).

The basis centrality measures like degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality have been studied in SNAM. Analysis, modeling, and interpretation of these centrality measures and their correlation has a great interest. In this study *Newman-Watts-Strogatz* and other social networks have been used to emphasize this interest.

The aim of this paper is to extract useful hidden structural information from social networks to maximize insight. For the degree, betweenness, closeness, and eigenvector centrality measures, the following are proposed: (i) An analysis and interpretation with an emphasis in the context of an infectious disease spread; (ii) A structural analysis and visualization of the Pearson correlation to emphasize the existence of a strong positive and structural or non-structural correlation; (iii) Some structural implications of centrality measures in a network.

The rest of this paper is organized as follows: the section 2 reviews the previous works, the section 3 presents the centrality measures study with interpretations and limits in area of *SNAM*. The methodology used is presented in the section 4, while the section 5 presents the empirical experimentations, results and discussion. We conclude in the section 6.

II RELATED WORK

The correlation between centrality measures and interpretation is studied from different perspectives in the literature. In [Valente et al. \(2008\)](#), the authors have shown the existence of a strong correlation between eigenvector centrality and degree centrality, with a mean correlation of 0.92 over 58 networks.

Moreover, the correlation between centrality metrics and their application to the opinion model is studied in [Li et al. \(2014\)](#). Authors have shown that betweenness, closeness centrality, and the components of the principal eigenvector of the adjacency matrix are strongly correlated with the degree, the 1st-order degree mass, and the 2nd-order degree mass, respectively, in both network models and real-world networks.

In [Meghanathan \(2016\)](#), author has examined the correlation analysis between maximal clique size and centrality measures for random networks and scale-free networks. Furthermore, [He and Meghanathan \(2016\)](#) have investigated the correlations between eigenvector centrality and five centrality measures including degree, betweenness, cluster coefficient, closeness and distance centrality in various types of networks. Their analysis shown that there was strong correlation between the degree centrality and the eigenvector centrality.

The rank correlation between centrality metrics in complex networks is studied in [Shao et al. \(2018\)](#). Authors have investigated the correlation between centrality metrics in real networks, and find that the betweenness occupies the highest coefficient, closeness is at the middle level, and eigenvector fluctuates dramatically.

However, with Data Science analysis and visualization technics, by extracting useful hidden information, we propose a structural analysis and visualization of the Pearson correlation to emphasize the existence of a strong positive structural or non-structural correlation between centrality measures study. Moreover, we show the limits of centrality measures, the stakes, and interpretation of these centrality measures in the context of infectious disease spread. Structural implications of these centrality measures on the network structure have been suggested.

III THE CENTRALITY MEASURES, INTERPRETATIONS AND LIMITS IN AREA OF SOCIAL NETWORK ANALYSIS

In network structural analysis, centrality measures are widely used to determine which nodes are important in a network. They constitute the structural importance of nodes in the network [Wasserman and Faust \(1994\)](#). The connectedness network is decisive for the usefulness of a centrality measure that is useful in the context of an infectious disease spread where the network is generally connected. In our application case of infectious disease spread, we are interested in the case of the non-oriented network. In the following definitions, we consider $G = (V, E)$ a graph/network by its adjacency matrix A , with n nodes and m links.

3.1 The degree centrality

The importance of a node is determined and increases with the number of its neighbors. For a node $v_i \in V$ it is defined by [Freeman \(1977\)](#): $C_d(v_i) = \sum_{j=1}^n (a_{ij})$; $C_d^*(v_i) = \frac{1}{n-1} \sum_{j=1}^n (a_{ij})$. Here $C_d^*(v_i)$ represents the standard notation. The degree centrality provides information only in the local popularity. This determines which nodes can quickly and directly propagated the virus in a strictly local area. Regarding time complexity analysis of the degree centrality, with A , we obtain $\theta(n^2)$. However, degree centrality becomes limited in a global vision of the network analysis.

3.2 The closeness centrality

This measure defines the global "close" distance of a node to other nodes [Freeman \(1977\)](#). For a node $v \in V$ it is defined by: $C_c(v_i) = \frac{1}{\sum_{t \in V} d_G(v_i, t)}$; $C_c^*(v_i) = \frac{n-1}{\sum_{t \in V} d_G(v_i, t)}$. Here $C_c^*(v_i)$ represents the standard notation, and $d_G(v, t) = d_G(t, v)$ the distance between two nodes v and t . This is important to further understand information dissemination or propagation. In a social network, this measure means that an node is important if he can easily reach a large number of person with a minimum size in the paths. However, some values the closeness centrality would be undefined since geodetic distances between some nodes would not exist. Regarding time complexity analysis of the closeness centrality, with the *Dijkstra* algorithm, we obtain $\theta(n^2)$. This complexity can be reduced to $\theta(n + m)$.

3.3 The betweenness centrality

In this concept a node is most important or influential since it is necessary to cross it and go from one node to another. For a node $v \in V$ it is defined by [Freeman \(1977\)](#):

$$C_b(v_i) = \sum_{s \neq v_i \neq t \in V} \frac{\sigma_{st}(v_i)}{\sigma_{st}}; C_b^*(v_i) = \frac{2}{(n-1)(n-2)} \sum_{s \neq v_i \neq t \in V} \frac{\sigma_{st}(v_i)}{\sigma_{st}}.$$

Here $C_b^*(v_i)$ is the standard notation $\sigma_{st}(v)$ is the total of shortest paths between nodes s, t that pass through nodes v , and σ_{st} is the total of shortest paths between nodes s and t . Nodes with higher score control the flow of transmission to impact the robustness in the network. These nodes play the role of bridges, hubs, critical links as well as cohesion between other nodes. Therefore, nodes with very high often join cohesive regions of the network, called communities. However, zero values of betweenness centrality can be observed in the case of a weakly connected network. Concerning time complexity analysis, it takes $\theta(n^3)$ time with the *Floyd-Warshall* algorithm. On a sparse graph, Johnson's algorithm taking $\theta(n^2 \log(n) + n * m)$, and even better, to $\theta(n * m)$ for unweighted graphs.

3.4 The eigenvector centrality

The eigenvector centrality provides global information that an influential node is connected to others well-connected nodes in a network [Bonacich \(2007\)](#). For a node $v_i \in V$ it is defined by: $\lambda C_e(v_i) = \sum_{j=1, i \neq j}^n a_{ji} C_e(v_j)$, where λ denotes an eigenvalue, a_{ji} is an entry of A , C_e is the eigenvector centrality. Assuming $x = (C_e(v_1), \dots, C_e(v_n))$; formal expression of the eigenvector centrality becomes: $Ax = \lambda x$ where computation implies to solve: $AC_e = \lambda C_e$. This measure presumes that not all connections have the same importance by taking into account not only the quantity but specifically the quality of these connections. Therefore, it represents the state towards which each node v_i evolves in the network. For j from 1 to n , each element a_{ij} is the state of the links or the evolution of interactions of one node with other nodes in the network. The eigenvector is a preferred direction unchanged by A , while λ represents the stretch or compression factor in that direction. However, the eigenvalues and eigenvectors of an antisymmetric matrix can be complex. The eigenvector centrality takes time complexity at least in polynomial time.

IV METHODOLOGY

We used the classical *SNAM* methodology in a process of six main orderly generic steps [Kazienko \(2018\)](#): (i) Problem definition; (ii) Data gathering and preparation; (iii) Social network modelling; (iv) Knowledge extraction; (v) Evaluation; (vi) Interpretation and deployment. We have enriched this with the use of descriptive, diagnostic, predictive, and prescriptive analysis. A specific *DataFrame* has been built.

The introduction of this *DataFrame* facilitates extraction of hidden useful information from networks in our dataset by structural analysis, representation, and interpretation. After around 100 simulations, cross-interpretation of data visualization results has been made to be able to focus on the structural correlation and interpretation between centrality measures studied.

V EXPERIMENTATIONS, RESULTS AND DISCUSSION

5.1 Empirical assumptions and experimental configurations

(i) Nodes and links are weakly organized at the global level; (ii) By definition, degree centrality is considered as an independent variable at the local level, while closeness, betweenness, and eigenvector centrality are dependent variables at the global level. An Intel core i7, 8GB RAM; the *Anaconda 3* software with the open-source *NetworkX* library was used and other data scientist libraries like *Numpy*, *Pandas*, *Scipy*, *Matplotlib*, and *Seaborn*.

5.2 Results and discussion

This study has been experimented on 31 social networks with different structural characteristics show in the figure 1 where we applied specific data analysis and visualization techniques in the dataset over around 100 experimentation series. Therefore, with the main *Newman Watts Strogatz* network, the figure 2 shows the front portrait and the figure 3 shows the heatmap diagram.

In the figure 3 "*dec*" means degree centrality, "*bsc*" betweenness centrality, "*clc*" closeness centrality and "*evc*" eigenvector centrality. In the figure 2 and figure 3 we observe a quantitative and qualitative values of very strong positive structural correlation between the four centrality measures study. In fact, betweenness, closeness, and eigenvector centrality depend to the degree

Id	Network	Type	Nodes	Links	kmax	k	k2	den	cc
0	Sociopatterns-infectious	Real social network	410	2765	50	13.488	252.434	0.032978	0.456
1	Airline	Real social network	36	57	20	3.167	33.389	0.090476	0.000
2	Internet	Real social network	40	61	10	3.050	13.000	0.078205	0.154
3	Karate club	Real social network	34	78	17	4.588	35.647	0.139037	0.571
4	Davis Southern Women	Real social network	32	89	14	5.562	39.062	0.179435	0.000
5	Florentine families	Real social network	15	20	6	2.667	8.933	0.190476	0.160
6	Les miserables	Real social network	77	254	36	6.597	79.532	0.086808	0.573
7	Watts Strogatz 1	Generated social network	1000	2000	10	4.000	17.898	0.004004	0.007
8	Watts Strogatz 2	Generated social network	3000	12000	18	8.000	67.741	0.002668	0.008
9	Connected Watts Strogatz 1	Generated social network	1000	2000	11	4.000	17.778	0.004004	0.008
10	Connected Watts Strogatz 2	Generated social network	3000	12000	16	8.000	68.055	0.002668	0.007
11	Newman Watts Strogatz 1	Generated social network	1000	3613	12	7.226	54.172	0.007233	0.157
12	Newman Watts Strogatz 2	Generated social network	5000	36028	23	14.411	211.592	0.002883	0.200
13	Newman Watts Strogatz 3	Generated social network	24	84	9	7.000	49.667	0.304348	0.498
14	Barabasi Albert	Generated social network	1000	4975	150	9.950	211.636	0.009960	0.044
15	Barbell	Generated social network	1005	1010	5	2.010	4.066	0.002002	0.005
16	Random 1	Generated network	1000	3500	17	7.000	56.174	0.007007	0.009
17	Random 2	Generated network	140	6811	111	97.300	9495.843	0.700000	0.700
18	Dense gnm Random	Generated network	1000	3500	16	7.000	56.016	0.007007	0.009
19	Random regular 1	Generated network	1000	1500	3	3.000	9.000	0.003003	0.003
20	Random regular 2	Generated network	1000	4500	9	9.000	81.000	0.009009	0.007
21	Facebook 1	Real social network	52	146	18	5.615	48.692	0.110106	0.462
22	Facebook 2	Real social network	61	270	29	8.852	109.705	0.147541	0.733
23	Facebook 3	Real social network	168	1656	77	19.714	645.321	0.118050	0.534
24	Facebook 4	Real social network	150	1693	57	22.573	680.240	0.151499	0.670
25	Facebook 5	Real social network	333	2519	77	15.129	469.526	0.045570	0.508
26	Facebook 6	Real social network	224	3192	99	28.500	1312.554	0.127803	0.544
27	Facebook 7	Real social network	534	4813	107	18.026	539.884	0.033820	0.544
28	Facebook 8	Real social network	786	14024	136	35.684	2086.852	0.045458	0.476
29	Facebook 9	Real social network	1034	26749	253	51.739	4886.236	0.050086	0.526
30	Facebook 10	Real social network	747	30025	293	80.388	10593.861	0.107759	0.635

Figure 1: The synthesis of the structural information about the networks in the dataset

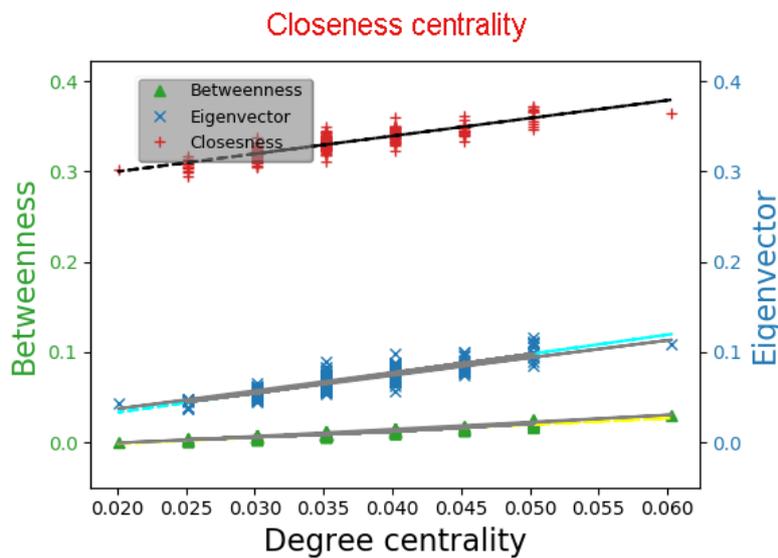


Figure 2: The front portrayal with linear regression between degree, closeness, betweenness and eigenvector centrality

centrality. Hence the local structure influence the global structures. These observations are similar to those in the literature [He and Meghanathan \(2016\)](#), [Li et al. \(2014\)](#), [Meghanathan \(2016\)](#).

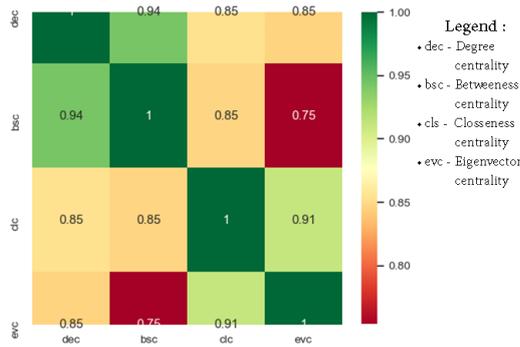


Figure 3: The heatmap diagram of the *Pearson* correlation between degree, closeness, betweenness and eigenvector centrality

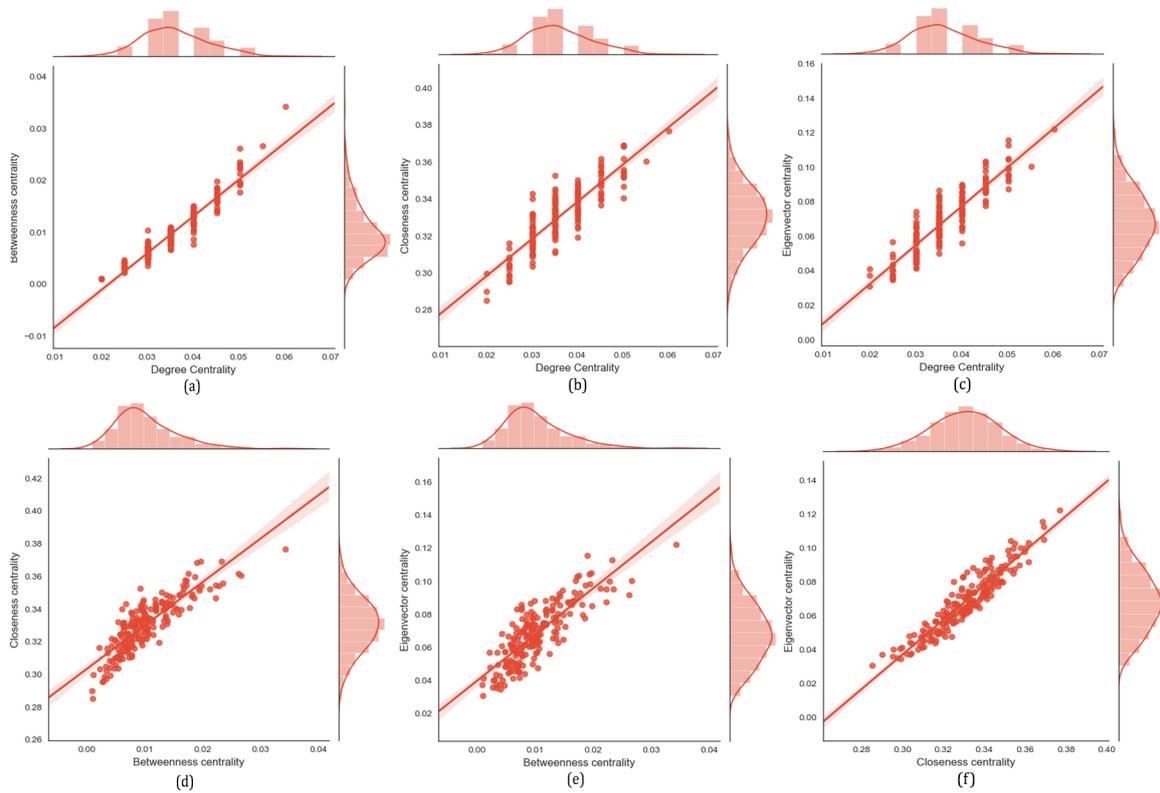


Figure 4: The scatter plot of centrality measures with linear regression for (a) degree and betweenness, (b) degree and closeness, (c) degree and eigenvector, (d) betweenness and closeness, (e) betweenness and eigenvector, (f) closeness and eigenvector

Overall the couples of centrality measure, the insight of the figure 4 show the existence of a very strong positive and structural correlation between degree and closeness centrality, degree and betweenness centrality, degree and eigenvector centrality. However, there is the existence of a strong positive and non-structural correlation between eigenvector and betweenness centrality, betweenness and closeness centrality, closeness and eigenvector centrality. The quality of the strong positive correlation between the pairs of centrality measures studied shows a strong positive structural dependence between these centrality measures. This correlation also reflects a similarity of relationships between centrality measures at the latent factors at structural levels ranging from local to global. However, the structure of correlations between centrality measures is not always orderly. These results confirm the hypothesis that the local structure strongly influence the global structures so that the global structures depend and emerge from the local

structures. With the same experimentation protocol, overall the 31 social networks at different topologies, the same results were observed. Moreover, in the figure 5 we propose that the *Venn diagram* can be used to understand the implication of the strong positive and structural correlation between centrality measures studied.

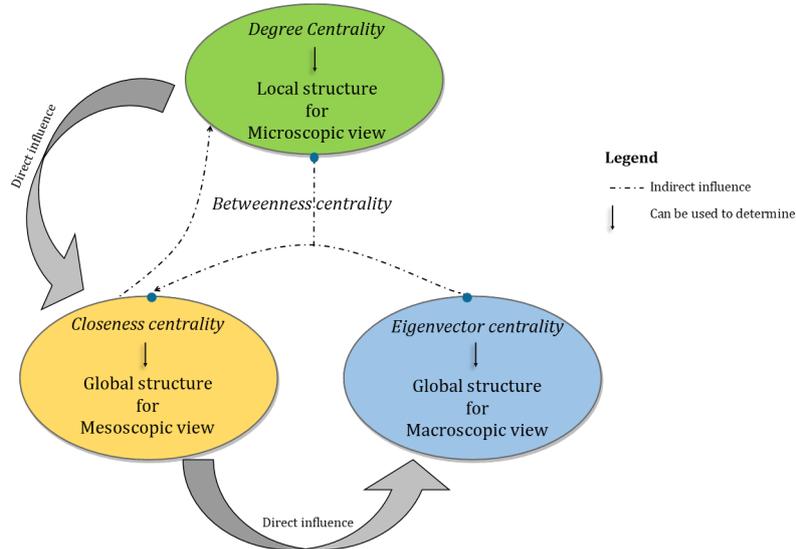


Figure 5: The implication of the correlation centrality measures in the network structure

With the works in [Daokun et al. \(2017\)](#), by the strong positive structural or non-structural correlation between degree, closeness, betweenness, and eigenvector centrality; we propose the figure 5 to categorize the structural implications of the centrality measures studies in the network structure. Therefore, it is argued that: the degree centrality influences the microscopic structure at the local proximity; closeness and betweenness centrality influence the mesoscopic structure as the structural role proximity and intracommunity proximity; eigenvector centrality influences the macroscopic structure at the network whole properties. Hence, we observed that the local structure has a great influence on the network global structure.

Identify influential nodes insight to the critical nodes from which a virus spreads with a high-spreading capacity in the network. This is useful to build an effective policy that mitigates infectious disease spread such as the priority vaccination, quarantine.

VI CONCLUSION

In this paper, we address the network structural correlation analysis problem using centrality measures in the context of the infectious disease spread. We focus on the correlation structures between centrality measures in the context of an infectious disease spread. We have studied the structural correlation between degree, closeness, betweenness, and eigenvector centrality. Here, network connectedness limits were established for these centrality measures, we suggest their issues, specific interpretations, and limits in the context of infectious disease spread. By extracting useful hidden structural information, we propose a structural analysis and visualization of the Pearson correlation between centrality measures study. The results show the existence of a strong positive and structural correlation between degree and closeness centrality, degree and betweenness centrality, degree and eigenvector centrality. However, a strong positive and non-structural correlation was observed between eigenvector and betweenness centrality, betweenness and closeness centrality, closeness and eigenvector centrality. This correlation shows

a strong positive structural dependence between these centrality measures that reflects a similarity of relationships between centrality measures study. However, the structure of correlations between centrality measures is not always orderly. For any node in the network, whatever the centrality measure, the higher is its score, the more important or structural influencer it becomes, the more it occupied a strategic propagation position on the network; and the more it increases the likelihood or risk of propagation. This can be used to build an effective policy that mitigates infectious disease spread. For the structural implications of these centrality measures, the degree centrality is related to the microscopic structure, closeness centrality and betweenness centrality to the mesoscopic structure, and eigenvector centrality to the macroscopic structure of the network. Further research attempts can be carried out to explore other structural properties in network distribution and segmentation. Also, we can experiment the gain insight with the use of the systemic approach to improving the general and the whole causal interpretation of the positive and structural correlation.

REFERENCES

- Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, 29.
- Danon, L., Ford, A., House, T., Jewell, C., Keeling, M., Roberts, G., Ross, J., and Vernon, M. (2011). Networks and the epidemiology of infectious disease. *Interdisciplinary perspectives on infectious diseases*, 2011:284909.
- Daokun, Z., Yin, J., Zhu, X., and Zhang, C. (2017). Network representation learning: A survey. *IEEE Transactions on Big Data*, PP.
- Enright, J. and Kao, R. (2018). Epidemics on dynamic networks. *Epidemics*, 24. ELSEVIER.
- Freeman, L. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40.
- He, X. and Meghanathan, N. (2016). Correlation of eigenvector centrality to other centrality measures : Random, small-world and real-world networks. *NeCoM, CSITEC*.
- Kazienko, P. (2018). Process of social network analysis. *Encyclopedia of Social Network Analysis and Mining*, pages 1927–1942.
- Keeling, M. and Rohani, P. (2008). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.
- Li, C., Li, Q., Mieghem, P., Stanley, H., and Wang, H. (2014). Correlation between centrality metrics and their application to the opinion model. *The European Physical Journal B*, 88.
- Meghanathan, N. (2016). Correlation analysis between maximal clique size and centrality metrics for random networks and scale-free networks. *Computer and Information Science*, 9:41.
- Pellis, L., Ball, F., Bansal, S., Eames, K., House, T., Isham, V., and Trapman, P. (2014). Eight challenges for network epidemic models. *Epidemics*, 10. ELSEVIER.
- Shao, C., Cui, P., Xun, P., Peng, Y., and Jiang, X. (2018). Rank correlation between centrality metrics in complex networks: An empirical study. *Open Physics*, 16:1009–1023.
- Valente, T., Coronges, K., Lakon, C., and Costenbader, E. (2008). How correlated are network centrality measures? *Connections (Toronto, Ont.)*, 28:16–26.
- Wasserman, S. and Faust, K. (1994). *Social Network Analysis : Methods and Applications (Structural Analysis in the Social Sciences)*. Cambridge University Press.