



The Density With Mutual Information In Communities-Based For Identifying The Nodes Influencing In Complex Weighted Networks

Hussein L. Hasan¹, Salah A. Albermany²

¹University of Babylon-Iraq

²University of Kufa-Iraq

Corresponding author: Hus_Loia@uobabylon.edu.iq

Abstract.

This paper presents a new method in identifying the influential and dominant nodes of information exchange within social networks. The procedure involves blending the characteristics of structured communities with exchanging information between nodes—the density with Mutual Information-Based Centrality (DMIBC). The results obtained from the method were compared, and a high-level correlation was found between standard metrics. DMIBC gave a new concept, which is not relying on network properties so that some nodes have a significant position within the community structure. Still, they tend to be idle and without interaction. Therefore, DMIBC gave high values to the nodes that have high rates of information exchange and at the same time have important positions within the community structure with the effect of density and number of components of the community in the composition of the method.

Keywords: Density, Mutual information, Communities, Influential nodes.

DOI Number: 10.14704/nq.2022.20.8.NQ44919

NeuroQuantology 2022; 20(8): 8997-9006

8997

INTRODUCTION

Many connections are created every day by new or prior social media users. Understanding the inner interactions of complex networks is critical in all areas of science [1] [2]. Billions of individuals are now weaving a diverse relations array of Yahoo, Twitter, WeChat, WhatsApp, Facebook, Instagram, Quick smartphone text messages, Exchanged Videos, Podcasts, Audio Streams of recordings, Articles, Wikis, Chat Forums, and Environments for virtual reality games to link them to individuals for whom

they care [3] [4]. In any of these platforms, a dynamic interplay takes place between social activities and technical infrastructures [5]. Working within any network to extract influencers gives wide use in the game theory, disease propagation, the management of rumors, the finding of social center characters, and the location of major transportation hubs [6] Within many networks, nodes differ in the amount of power and influence they have [7] [8]. The emphasis has moved from focusing on one person communicating with a virtual



system to growing social engagement, experience, and mutual intellect as the environment has become progressively linked [9]. Individuals gradually use mobile devices that can connect content to access these instruments in real-time to locations [10]. Technologies that facilitate social networking are some of the most relevant. The miracles of our time, the extraordinary rise and growth of social mediation technology usage, have fundamentally created fresh ways to work, play, and build meaning [11] [12]. It is leaving an indelible imprint on almost any possible domain. Given the rising ubiquity of social media, it has hardly been feasible to harness their capacity [13]. Using and developing social technology successfully is highly valued. The links in a social network form communities and bonds that are described as social relations [11] [14]. A relationship may start with an unimportant associate and may be communicated in numerous ways [15] [16]; it may develop over time and lead to important relationships. Analysts often do understand networks to provide solutions about extracting important nodes, locations of subgroups, or getting a view of how a network topology relates to other networks. This is done through network visualization or by using metrics developed by researchers.

Working within any network to extract influencers without regard to the communities it contains is a futile act. The individual is not active in all the joints of the network but rather is active within the interested communities and interacts with them.

Traditionally, nodes significance ranking measurement indices based on the structure of the network were derived from locally and globally attributes of the network, the location of the nodes, and randomized walks to arrive at their values. While currently, strong methods are being adopted to reach the communities within these networks and show the influencers in those networks. Many metrics can be used, but the problem is to choose the appropriate metric that reflects the true image of

influencers [17] [18]. Therefore, the metrics can be divided into several sections, depending on the vision of the network. Where the network element can be viewed based on three different levels:

- Dyadic
- Nodal
- Group

On the basis of it, the metrics or properties that deal with each part are dealt with.

Relational data is the foundation of every social network study. Particularly, network data that captures the social relationship between the two components. (i.e., a dyad). Each of these types of relational data may be either undirected or directed; they could be categorical, valued, or binary; they are either positive or negative [19]; they count interactions or flow of information at a dyad [7]. Through the dyadic level, it can extract information about values, norms, and the behavior of transmissions between nodes.

Although dyadic social relationships serve as the fundamental parts of any network analysis, the analysis itself may take place at a variety of different levels. The three most often used levels of analysis are dyadic, nodal (or individual), and group [20][21].

Network measures such as network complexity, density, and relatedness enable researchers to examine systematically and detect the patterns of connection inside social networks, establish a basis for comparing networks, follow alterations in the social networks throughout time, and evaluate people's and groups' relative positions inside a social network [22].

Initially, social network metrics concentrated on the most fundamental counts of connections, such as the number of friends or followers, and then eventually expanded into more complex calculation methods as the social network expanded and encompassed ideas of network density, transitivity, centrality, balance, and structural holes [23] [24].



While some metrics reflect the network as a whole, many others characterize subsets of the network. An example of this is the nodes count, which represents the network's vertex count, whereas the links count indicates a quantity of edges among them. There is the other entity network metric called "density" that quantifies the degree to which a collection of nodes is connected. This density metric calculates the proportion of links assuming the maximum density of each node connected to the entire network.

Additional measurements are made at a single node in the social network; as examples, there are several measurements of "centrality." Which it finds ways to quantitatively assess how "essential" (of importance) a vertex is inside the network and use it to guide design decisions. Some pairs of persons are not linked directly in several human networks, including highly linked networks. If a third member connects ("friend of a friend"), this member may be considered a "connector," "bridge," If the individual is

$$Ad_G = \frac{\sum_{i \neq j} Dis(s_i, t_j)}{n(n-1)} \quad (1)$$

Where: Dis(s,t) is shortest path between {s,t}, n is number of nodes in the network.

- **Network density**

The density of a network (DenG) is calculated by dividing the observed connections

$$Den_G = \frac{2m}{n(n-1)} \quad (2)$$

Where m is observed edges, n is number of nodes.

It is a quantifiable method of capturing significant social concepts like as membership, solidarity, and cohesiveness [29]. Centralization is a statistic that defines how much the network focuses on a single or a few significant vertices. Networks that are centralized feature numerous links branching off from a few key vertices, while decentralized networks have several vertices with many interactions [30].

- **Centrality of degree**

missing, the gap might be seen as a "structural hole" [25].

PRELIMINARIES

In the presence of a graph $G=(V,E)$, wherein V denotes vertices, E denoted links between vertices, and most essential metrics, which are average of distance (AdG), network density (DenG), centrality of betweenness (BC), centrality of degree (DC), centrality of closeness (CC). This section will provide an overview of the previous metrics.

- **Average of distance**

This is the average length of the distances between all pairs of nodes. It reveals whether network community members are close to one another. It is among the most significant considerations for improving networks for performance and power [26]. The distance Dis(s,t) between {s, t}, where $s, t \in V(G)$ can actually be considered as the shortest path between the nodes, where the average of distance (AdG) calculate by Equation (1) [27]:

by the total number of possible connections. Where Equation (2) calculate the density for the undirected graph [13] [28]:

There are many ways of comparing nodes' levels of centrality within a network. The most straightforward – degree centrality – is to compare their respective numbers of ties. The centrality of a degree is means the number of connections that link it. It's a form of popularity metric, but it's a rudimentary one that doesn't distinguish both quantity and quality. In addition, for every one of the various node graph metrics, a frequency plot is constructed. When examining vast networks, the frequency plots are quite useful. Mathematical expression for the centrality of degree in Equation [30] (3):



$$D. c(v_i) = \sum_{j=1}^{N_v} k_{ij} \quad (3)$$

Where k_{ij} denoted the link between node i and j .

- **Centrality of betweenness**

The betweenness centrality of node v is the portion of geodesic paths between other nodes that v falls on. Mathematical expression for Betweenness centrality in Equation (4) [31]:

$$B. c(v) = \sum_{s \neq t \neq v} \frac{Dis_{st}(v)}{Dis_{st}} \quad (4)$$

Where:

- Dis_{st} denoted the total of distance between node s and t .
- Dis_{st} (v) denoted the total distance between node s and t that pass through a node v .

- **Centrality of closeness**

Closeness centrality is defined as the inverse of the sum of the shortest distances between each node and every other node in the network. Mathematically, it can be expressed by the Equation (5):

$$C. c(v) = \frac{1}{\sum_{s=1}^N Dis(v,s)} \quad (5)$$

Where: Dis (v, s) denoted the distance between v & s . [31][8].

METHODOLOGY

This paper aims to find the right influence within communities, where there is always a tendency for individuals to interact within specific communities, and these communities differ in the intensity of relationships and the number of individuals.

However, we suppose that there are two behaviours of the node: the first, there is a strong connection and strong relations with its community, and the second behaviour is that there is a weak connection of the node with the rest of the communities.

A method for forming the technique depends on four essential sides

- First, the exchange between the vertices may include exchanging messages, pictures,

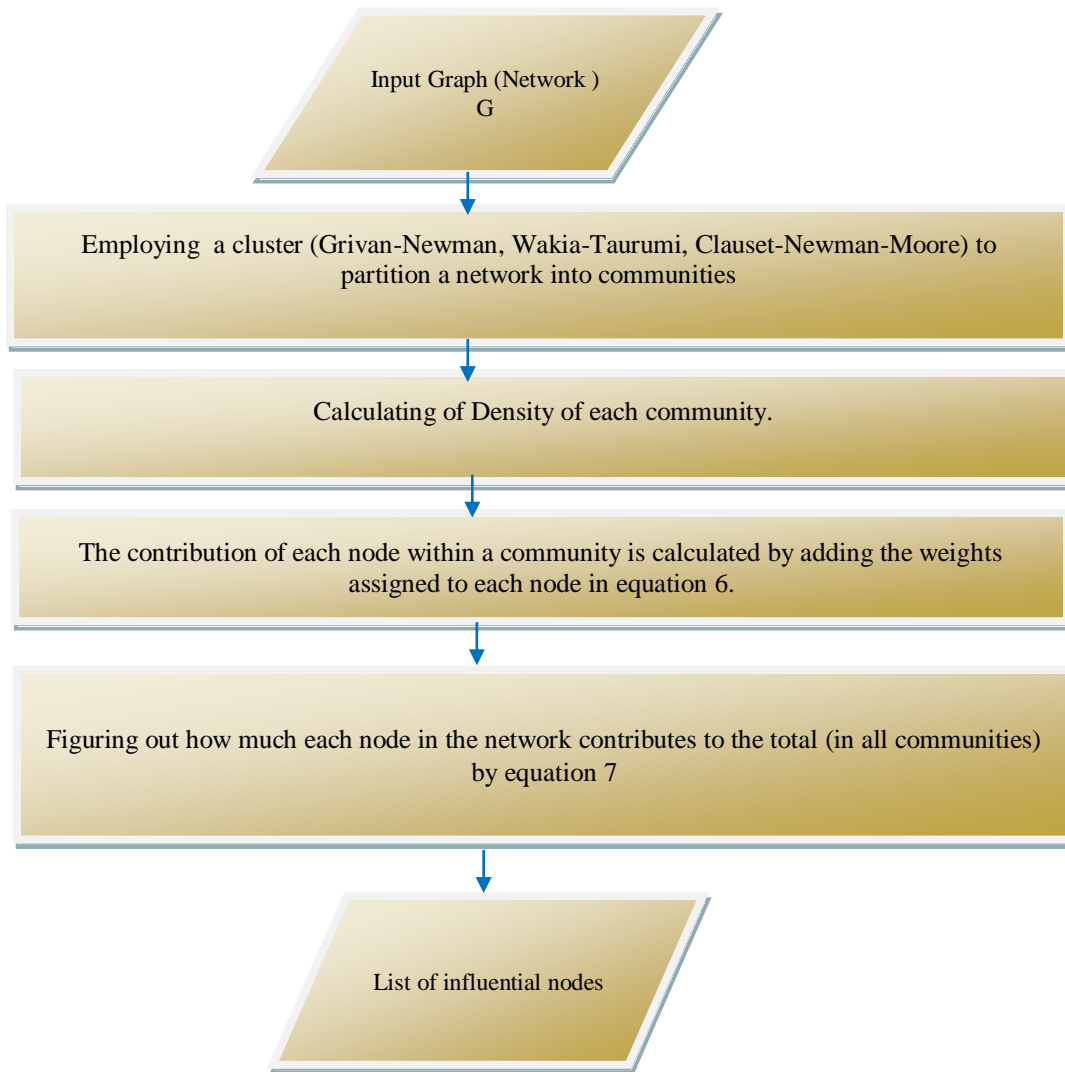
information, news, and others. Increased exchange means an excessive activity of the node within the community. Lead to the increased influence of that node

9000

- Second, the location of the node, which has a role in increasing the exchange between the node and its vicinity
- Third, the density of the community to which the node is connected, and the greater the density, the greater the effect of vertex.
- Fourth, the number of community components; increasing the number means ensuring access to the largest audience.

Figure. 1 showing the stages of (DMIBC) .





9001

Figure. 1 .The general outline ofDMIBC

Stages of (DMIBC)

- A. Input: undirected graph $g=(v,e)$
- B. Applying cluster to divided network into communities (Grivan-Newman, Wakia-Taurumi, Clauset-Newman-Moore)
- C. Calculating of Density of each community by apply Equation (2).
- D. Applying Equation (6). that calculate contribution of each node within each community by summation of weights for specific node.

$$\forall u \in \mathbf{k} \mathbf{W}_{uk} = \sum_{i=1}^D \mathbf{S}_{ui}, \mathbf{S}_{ui} \in \mathbf{V}(\mathbf{k})(6)$$

Where u is index of node inside of community k , \mathbf{S}_{ui} represent of weight between node u and

any node in community k , D is number of links of node u inside community k .



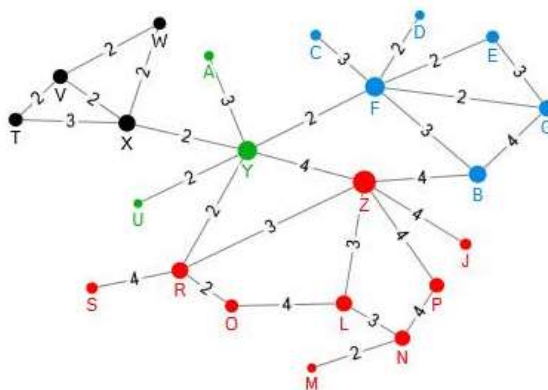
E. Calculating the contributions for each nodes on whole network (in all communities)

$$\forall u \in \mathbf{GDMIBC}_u = \sum_{k=1}^C W_{uk} \cdot \frac{Din_k \cdot n_k}{Din_G \cdot N_G} \quad (7)$$

Where u is node inside of graph G , W_{uk} represent of summation weights of node u in community k , C is number of communities in graph G , Din_k is density of community k , Din_G is density of graph G , n_k is represent of nodes number in community k , N_G is represent of nodes number in graph G .

Apply DMIBC on sample network

Using a simple network to show the algorithm results in detail, Figure. 2. shows the network as it consists of 22 nodes and 29 edges.



9002

FIGURE. 1. Sample network with 22 nodes &29 edges

Implementation begins with application of the Grivan-Newman to extract the communities & calculating of density of each community by apply Equation (2) as shown in Table 1.

Table 1 The communities & density of sample network

| Community No. | nodes inside community | Community Density |
|---------------|---------------------------|-------------------|
| 1 | M, N, P, L, J, Z, S, R, O | 0.277778 |
| 2 | U, Y, A | 0.666667 |
| 3 | V, W, X, T | 0.833333 |
| 4 | B, C, F, G, E, D | 0.466667 |

Calculating the contribution of each node within each community by summation of weights for specific node by applying equation (6). As shown in Table 2.

Table 2 The weights of nodes inside the communities

| Vertex | Community No.1 | Community No.2 | Community No.3 | Community No.4 |
|--------|----------------|----------------|----------------|----------------|
| M | 2 | 0 | 0 | 0 |
| N | 9 | 0 | 0 | 0 |



| | | | | |
|---|----|---|---|---|
| P | 8 | 0 | 0 | 0 |
| L | 10 | 0 | 0 | 0 |
| J | 4 | 0 | 0 | 0 |
| Z | 14 | 4 | 0 | 2 |
| R | 9 | 2 | 0 | 0 |
| S | 4 | 0 | 0 | 0 |
| O | 6 | 0 | 0 | 0 |
| U | 0 | 2 | 0 | 0 |
| A | 0 | 3 | 0 | 0 |
| Y | 6 | 5 | 2 | 2 |
| V | 0 | 0 | 6 | 0 |
| W | 0 | 0 | 4 | 0 |
| T | 0 | 0 | 5 | 0 |
| X | 0 | 2 | 7 | 0 |
| B | 4 | 0 | 0 | 7 |
| C | 0 | 0 | 0 | 3 |
| F | 0 | 2 | 0 | 9 |
| G | 0 | 0 | 0 | 9 |
| E | 0 | 0 | 0 | 5 |
| D | 0 | 0 | 0 | 2 |

The next steps is calculating the contributions for each nodes on whole network (in all communities)

9003

Node Y as example :

Table 3 displays links number & weights summation of node Y in each community :

Table 3. The links & weights of node Y

| Community No. | Links Number of node Y in each community | Weights summation of node Y in each community |
|---------------|--|---|
| 1 | 2 | 6 |
| 2 | 2 | 5 |
| 3 | 1 | 2 |
| 4 | 1 | 2 |

Apply Equation.7 on whole graph for node Y:

$$DMIBC_Y = ((6 * (0.277778 * 9) / (0.125541126 * 22)) + (5 * (0.666667 * 3) / (0.125541126 * 22)) + (2 * (0.833333 * 4) / (0.125541126 * 22)) + (2 * (0.466667 * 6) / (0.125541126 * 22))) = \mathbf{13.49311}$$

Table 4 shows the implementation of an Equation.7 for each node

Table 4. DMIBC values of all nodes

| Node | DMIBC | Node | DMIBC |
|------|-------------|------|-------------|
| Z | 19.62415755 | T | 6.034488795 |
| F | 13.61380671 | O | 5.431039914 |



| | | | |
|---|-------------|---|-------------|
| Y | 13.49311694 | E | 5.068970585 |
| B | 10.7172521 | W | 4.827591036 |
| X | 9.896561623 | J | 3.620693276 |
| R | 9.594837181 | S | 3.620693276 |
| G | 9.124147053 | C | 3.041382351 |
| L | 9.05173319 | A | 2.172415965 |
| N | 8.146559871 | D | 2.027588234 |
| V | 7.241386554 | M | 1.810346638 |
| P | 7.241386552 | U | 1.44827731 |

The results DMIBC compared with the centrality metrics

The importance of metrics in extracting network characteristics and applications was mentioned. The Katharina A. [1] clarified several points related to applying standard or centrality metrics. These metrics can be applied in various models to explain a network's characteristics. The following are the most important general outlines of the curricula that have been exploited metrics:

1. The Centrality metrics were employed to describe the network characteristics and overall structure.

2. The Centrality metrics are used to test the correlation with external criteria of the node or to test a model based on an analytical basis.

3. Directly metric-based operations can be configured to predict influential nodes within the network.

Table 5. shows a clear appearance of most of the values of the five metrics in the results of the DMIBC. This generates the conclusion that the DMIBC was able to a large extent to reconcile the five measures. So, its results reflected a clear diversity and comprehensiveness of a local and global scale on the network.

9004

Table 5The values of the metric standards with DMIBCfor sample network.

| Node | Degree | Betweenness | Closeness | Eigenvector | PageRank | DMIBC |
|------|--------|-------------|-----------|-------------|----------|-------------|
| Z | 6 | 87 | 0.023 | 0.105 | 2.111 | 19.62415755 |
| F | 6 | 65.333 | 0.020 | 0.096 | 2.158 | 13.61380671 |
| Y | 6 | 123.333 | 0.026 | 0.109 | 2.139 | 13.49311694 |
| B | 3 | 20.667 | 0.019 | 0.073 | 1.061 | 10.7172521 |
| X | 4 | 54.5 | 0.019 | 0.05 | 1.409 | 9.896561623 |
| R | 4 | 33.5 | 0.020 | 0.076 | 1.481 | 9.594837181 |
| G | 3 | 2.5 | 0.016 | 0.06 | 1.08 | 9.124147053 |
| L | 3 | 22.167 | 0.018 | 0.046 | 1.133 | 9.05173319 |
| N | 3 | 20.833 | 0.014 | 0.025 | 1.236 | 8.146559871 |
| V | 3 | 0.5 | 0.014 | 0.026 | 1.095 | 7.241386554 |
| P | 2 | 16.333 | 0.017 | 0.037 | 0.799 | 7.241386552 |
| T | 2 | 0 | 0.014 | 0.022 | 0.76 | 6.034488795 |
| O | 2 | 2.333 | 0.016 | 0.034 | 0.786 | 5.431039914 |
| E | 2 | 0 | 0.015 | 0.044 | 0.762 | 5.068970585 |
| W | 2 | 0 | 0.014 | 0.022 | 0.76 | 4.827591036 |
| J | 1 | 0 | 0.016 | 0.029 | 0.449 | 3.620693276 |
| S | 1 | 0 | 0.015 | 0.021 | 0.465 | 3.620693276 |
| C | 1 | 0 | 0.014 | 0.027 | 0.456 | 3.041382351 |
| A | 1 | 0 | 0.017 | 0.031 | 0.453 | 2.172415965 |



| | | | | | | |
|---|---|---|-------|-------|-------|-------------|
| D | 1 | 0 | 0.014 | 0.027 | 0.456 | 2.027588234 |
| M | 1 | 0 | 0.011 | 0.007 | 0.5 | 1.810346638 |
| U | 1 | 0 | 0.017 | 0.031 | 0.453 | 1.44827731 |

While Table 6 shows values correlation coefficient between standards metric with DMIBC.

Table 6 Results of correlation coefficient between DMIBC& standards metric

| | | | | |
|----------|-------------|----------|-------------|----------|
| PageRank | Eigenvector | Degree | Betweenness | PageRank |
| 0.938411 | 0.877322 | 0.946453 | 0.828225 | 0.776399 |

CONCLUSION

The proposed method gave a wide variety of network properties, where DMIBC assumed that within each network there should be a set of divisions of communities. Within these societies, the individual interacts in different proportions depending on the interest. The exchange between the vertices may include exchanging messages, pictures, information, news, and others. Increased exchange means an excessive activity of the node within the community. Lead to the increased influence of that node. The DMIBC gave an important role to the location of the node within the network in highlighting its importance. The important principle that the DMIBC relied on is the density of the communities within which the interaction takes place. The DMIBC also took into account the number of nodes within each community. Through these features, the DMIBC was able to succeed in reconciling the local and global features of the network through the results of the correlation and comparing with standard metrics. All this gave preference to extracting influencers within their communities in which they interact and the network.

References

[1] N. A. and C. S.H., Agent-based modeling and network dynamics, Oxford University Press, 2016.
 [2] E. Gilbert and K. K. , "Predicting tie strength with social media," in *Proceedings of the SIGCHI conference on human factors in*

computing systems (pp. 211-220), 2009.

[3] D. N. T. S. Agarwal N, Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining, Springer, 2018.
 [4] N. ME, "The structure and function of complex networks," *SIAM review*, vol. 2, pp. 167-256, 2003.
 [5] R. a. R. J. Alhaji, Encyclopedia of social network analysis and mining, Springer, 2020.
 [6] R. a. J. R. Alhaji, Encyclopedia of social network analysis and mining, Springer, 2014.
 [7] S. P. M. A. B. D. & L. G. Borgatti, "Network analysis in the social sciences," *Science*, vol. 5916, no. 323, p. 892–895, 2009.
 [8] M. Kaya, J. Kawash, S. Khoury and M. Y. Day, Social Network Based Big Data Analysis and Applications., Springer., 2018.
 [9] S. H. Bornholdt S, Handbook of graphs and networks, From Genome to the Internet, Willey-VCH , 2001.
 [10] Š. L. B. M. Blagus N, "Self-similar scaling of density in complex real-world networks," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 8, pp. 2794-802., 2012.
 [11] G. J., Analyzing the Social Web, Morgan Kaufmann . ImprintofElsevier, 2013.
 [12] H. D. a. S. I. Shneiderman B, Analyzing Social Media Networks with NodeXL., Morgan Kaufmann. imprint of Elsevier,



- 2020.
- [13] S. R. “. J. F. Peters, "Proximal three-way decisions: theory and applications in social networks," *Knowledge-Based Syst., Elsevier*, vol. 91, no. 1, p. 4–15, 2016.
- [14] H. & A. S. Hasan, "The Dominant Nodes Index in Shortest Paths (DNISP) as a Local Measure of Source Looking for the Target in Complex Networks," *International Journal of Interactive Mobile Technologies (IJIM)*, vol. 14, no. 19, pp. 97-108, 2020.
- [15] L. M. B. R. Zhang L, "Product information diffusion in a social network," *Electronic Commerce Research*, vol. 1, pp. 3-19, 2020.
- [16] M. S. S. C. a. A. H. Paidar, "Identifying Top-k Most Influential Nodes by using the Topological Diffusion Models in the Complex Networks.," *Network*, vol. 5, pp. 1-10, 2017.
- [17] F. R. S. M. C. N. Molaei S, "Identifying influential nodes in heterogeneous networks," *Expert Systems with Applications*, vol. 160, p. 113580, 2020.
- [18] Z. KA., *Network Analysis Literacy_ A Practical Approach to the Analysis of Networks*, Springer, 2016.
- [19] N. M. L. G. J. & A. F. Harrigan, "Negative ties and signed graphs research: Stimulating research on dissociative forces in social networks," *Social Networks*, vol. 1, no. 60, pp. 1-10., 2020.
- [20] P. V. Marsden, "Network data and measurement," *Annual Review of Sociology*, vol. 16, no. 1, p. 435–463, 1990.
- [21] N. S. W. S. & F. K. Contractor, " Testing multitheoretical, multilevel hypotheses about organizational networks: An analytic framework and empirical example," *Academy of Management Review*, vol. 31, no. 3, p. 681–703, 2006.
- [22] M. S. & M. C. Mizruchi, "Egocentric, sociocentric, or dyadic? Identifying the appropriate level of analysis in the study of organizational networks," *Social Networks*, vol. 28, no. 3, p. 187–208, 2006.
- [23] D. J. & B. S. P. Brass, *A practical guide to multilevel social network research, the handbook of multilevel theory, measurement, and analysis*, American Psychological Association, 2019.
- [24] D. Hansen, M. A. S. B. Shneiderman and I. Himelboim, *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*, Morgan Kaufmann, 2019.
- [25] L. J. B. M. e. Fu X, *Social network analysis: interdisciplinary approaches and case studies*, CRC Press, 2017.
- [26] M. E. Newman, "The mathematics of networks," *The new palgrave encyclopedia of economics*, vol. 2008, no. 2, pp. 1-12, 2008.
- [27] B. Bollobás, *Random Graphs*, New York: Academic Press, 2nd, 2001.
- [28] M. NEWMAN, *Networks*, Oxford university press, 2018.
- [29] J. a. B. J. Podolny, "Resources and relationships," *American Sociological Review* , vol. 62, no. 1, p. 673–693, 1997.
- [30] L. Freeman, "Centrality in social networks: conceptual clarification," *Social Network*, vol. 1, no. 12, pp. 35-41, 1979.
- [31] P. Bonacich, "Power and centrality: A family of measures," *American Journal of Sociology*, vol. 92, no. 5, p. 1170–1182, 1987.

