Journal of Theoretical and Applied Information Technology
15th January 2024. Vol.102. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IDENTIFYING INFLUENTIAL NODES IN DIRECTED WEIGHTED NETWORKS USING PYTHAGOREAN FUZZY SETS

VENKATA RAO SONGA[1] , DR. PRAJNA BODAPATI[2]

[1]Research Scholar, Department of CS & SE, University College of Engineering , Andhra University, India

[2]Professor, Department of CS & SE, University College of Engineering , Andhra University, India

E-mail: [1]songavenkat@gmail.com, [2]prof.bprajna@andhrauniversity.edu.in

ABSTRACT

Centrality considers node importance in complex networks, addressing this issue poses a significant challenge in the realm of social network analysis. Over the recent years, various measures of centrality have been suggested to evaluate the impact of nodes inside a network. However, these measures have certain drawbacks based on the network structure , lack of ground-truth values etc. This study introduces a new centrality metric called Node Pack Fuzzy Information Centrality (NPFIC), which suggests that crucial information about a node's significance can be derived from the internal structure of its pack. NPFIC quantifies the significance of a node by assessing the information content within its pack, which is calculated by the improved Havrda and Charavat entropy. We use Pythagorean Fuzzy Sets to address the uncertainty associated with the contributions of neighboring nodes to the centrality of the center node, this is often overlooked by established traditional approaches. To illustrate the effectiveness of the proposed approach, we compare it with four established centrality measures. We conduct experiments on a real-world directed weighted complex network to validate its performance and we employed the susceptible-infected-recovered (SIR) model to assess the effectiveness of our proposed approach. The outcomes of our experiments reveal that the crucial nodes identified by NPFIC significantly influence network connectivity.

Keywords: *Directed Weighted Complex Networks, Pythagorean Fuzzy Sets (PFSs), Node Pack Fuzzy Information Centrality(NPFIC), SIR , Havrda-Charvat Entropy, Centrality Measures.*

1. INTRODUCTION

In recent studies, numerous real-world systems are effectively modeled as complex networks, including social networks, the internet, power grids, and various online networks that significantly impact our daily lives. The analysis of complex networks[1] enables us to comprehend the complexities of unpredictability and forecast the evolution of systems. Within these complex networks, specific nodes exhibit notable influence, and have great impact on the overall structure and functionality. If a node has greater influence, then it plays crucial roles in facilitating information exchange, and their removal can substantially alter the network dynamics. Consequently, the identification of these influential nodes remains a focal point in the ongoing research on complex networks.

Numerous centrality metrics predominantly address unweighted undirected networks, often overlooking the significance of edge weights and directions. When dealing with directed networks, the weights of connections between the nodes can hold crucial information and should be considered. In many real-world scenarios, the strength, intensity, or some other quantitative measures associated with the relationship between nodes can greatly influence the network's behavior and characteristics. Directed weighted networks offer a more effective means of accurately depicting the relationships among individual nodes compared to simple networks. The inclusion of edge weight information proves invaluable in gaining a more precise understanding of the structure and functionality of the network. Consequently, detecting crucial nodes in networks with directed weights holds significant promise in research domains such as the propagation of influence in social networks, strategic planning for transportation infrastructure, disease spread and epidemiology, as well as ranking in scientific collaborations among others. The centrality metrics employed in complex networks with directed and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

weighted edges can effortlessly extend beyond those originally formulated for simple networks.

L. Zadeh introduced Fuzzy Sets (FSs)[2], a concept widely employed across various domains, including uncertainty measurement[3], multiple attributes decision making[4], analyzing and supporting decision-making[5], information granules[6], and similarity measures[7]. Unlike Boolean logic, Fuzzy Sets excel at quantifying uncertain information, making them well-suited for complex networks. Pythagorean membership grades for multicriteria decision making were first proposed by Yager[8]. Zhang and Xu [9] expanded the application of the Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) to incorporate Pythagorean and hesitant fuzzy sets in the context of multiple criteria decision-making. More recently, Xue et al.[10] presented the Pythagorean fuzzy LINMAP method, incorporating entropy for effective decision making in railway project investments. Shannon entropy, a concept in information theory, plays a crucial role in evaluating the anticipated information within a message. Its applications extend to the analysis of complex networks, particularly in identifying influential spreaders[11]. Zareie et al.[12] introduced the Entropy-Based Ranking Measure (ERM), emphasizing the concept that influential spreaders are characterized by substantial and evenly distributed degrees. ERM assesses the entropy of the immediate neighbors of a node those at the second order and concentrating on local information. However, it exclusively considers the node's immediate and secondary neighbors. Deng and Wen[13] applied Shannon entropy in their study to evaluate node importance through the introduction of the LID model. Even though LID employs Shannon entropy for quantifying information within a node box, it does not investigate the internal structure of these boxes. Traditional centrality metrics, including Degree Centrality(DC) [14], Betweenness Centrality (BC) [15], Closeness Centrality (CC) [16], Eigenvector Centrality (EC) [17], and PageRank (PR) [18], have been developed to assess the importance of nodes in a network by considering the factors like node distance and the number of connections. These traditional approaches focus on various characteristics of complex networks.

Wang et al.[19] investigated the detection of pivotal nodes within directed biological networks, relying on the characterization of node importance derived from instances observed in diverse networks with 2, 3, and a subset of 4 nodes. Additionally, Sheng et al.[20] introduced the concept of influential nodes in complex networks. Wang et al.[21] introduced a novel approach for discerning influential nodes in complex networks through a semi-local measure. Panfeng[22], implemented a voting methodology to recognize pivotal nodes within social networks. CC is only applicable to undirected networks, while PR and EC are generally used for directed networks. Garas and colleagues [23] presented a technique for k-shell decomposition in weighted networks. It can identify the nodes with the greatest impact in a network with weights by splitting a network into k-shell structure. PageRank asserts that a node's significance in web page ranking is tied to the quantity and quality demonstrated by the neighboring nodes.

However, these approaches overlook the wider configuration of networks. In addressing this issue, Betweenness Centrality (BC) examines the centrality of a node by assessing the number of shortest paths that traverse through it. On the other hand, Closeness Centrality (CC) posits that a node with the minimal average distance to other nodes wields greater influence. Although BC and CC prove to be effective, they are hindered by computational complexity, leading to suboptimal performance in complex networks.

In our discussions so far with different approaches [12]–[23] one problem identified is that all the models rely on observing the entire network and due to computational complexity, implementing this approach becomes impractical for large-scale social networks. Additionally, the center node receives the most significant contributions from its neighboring nods and their edge weights, are considered as an important concept in the directed weighted networks. It is a very typical task to identify vital nodes by considering all these features of individual nodes in directed weighted networks. Many current centrality metrics only consider the overall count of neighbors connected to a node and neglect an in-depth analysis of the local network structure, leading to inaccuracies.

Here, we introduce NPFIC as an approach to assess the significance of nodes within directed weighted networks by considering two factors: in-weights and out-weights. Based on the above discussions, the motivating factors behind our study can be succinctly summarized as follows:

i) to account for the variability in centrality values within a specific node due to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

uncertainty in edges, utilize Pythagorean fuzzy relations.

ii) to explore and analyze the inner structure of a node's pack within directed weighted networks and ascertain the amount of information it encompasses.

iii) to address the uncertainty linked with contributions from neighboring nodes to their central node, by leveraging Pythagorean fuzzy sets.

iv) to enhance the efficiency of Havrda and Charavat entropy and to optimize its utilization in directed weighted networks.

The main objectives of this article can be outlined as follows:

i) NPFIC mainly focuses on the internal arrangement of a node's pack to determine its global significance, this involves treating a node's pack as a Pythagorean set and using an improved Havrda-Charavat entropy for assessing the level of certainty within the pack. This method provides enhanced accuracy.

ii) In contrast to existing methods, NPFIC recognizes that the influence of neighboring nodes on a central node's importance is closely tied to their degree of membership and non-membership, so it employs fuzzy sets to distribute different weights to neighbor nodes within a node's pack with appropriate considerations towards directed weighted networks.

iii) NPFIC measures and employs the improved Havrda-Charavat entropy to assess the node's importance and its ability to spread in directed weighted networks.

This paper is organized in the following manner: Initially, we provide a concise overview of existing centrality measures. Following that, we articulate our method for identifying influential nodes, encompassing the calculation of shortest distances between two vertices in a directed network with assigned weights and the application of Havrda-Charvat Entropy. Subsequent sections include a comparative analysis of experimental results and analysis. Lastly, we present our conclusions.

2. PRELIMINARIES

2.1 Definition Of Pythagorean Fuzzy Set

Consider a universal set X. A Pythagorean fuzzy set (PFS)[24], denoted as P on X is defined as $P = \{\langle x, \mathrm{Mp}(x), \mathrm{Np}(x)\rangle : x \in X\}$ where $M_P(x):X\rightarrow[0,1]$ represents the membership degree, and $N_P(x):X\rightarrow[0,1]$ represents the non-membership degree of x∈X to the set P. This is subject to the condition $0 \le (M_P^2(x) + N_P^2(x)) \le 1$. For simplicity, Zhang, and Xu[9] introduced the notation $(M_P(X),N_P(X))$ as a Pythagorean fuzzy number (PFN), denoted as $P=(M_P,N_P)$. For any PFS P in X, the value $I_P(X)$ called the Pythagorean index of the element x in P, representing the indeterminacy or hesitancy of an element x∈X.

2.2 Centrality Measures in a Graph

Centrality measures utilize the graph's topological configuration to evaluate the importance of individual nodes. Local indicators, like degree centrality, focus on the intrinsic characteristics of each node, while Semi-local measures such as the h-index [25] and entropy centrality [26] analyze node connections to determine its significance. Conversely, global metrics like closeness centrality [27] and k-shell centrality [28] necessitate a thorough exploration of the entire graph to assess the centrality of each node.

2.2.1 Degree Centrality(DC):

DC provides the most straightforward measure for depicting the importance of nodes in networks with weights. In weighted networks, the degree centrality considers two important parameters which are strength and degree of the node. Opsahl et al.[29] introduced a generalized degree centrality for weighted networks, incorporating these parameters. This centrality measure is defined as:

$$C_D^{w\alpha}(i) = k_i^{1-\alpha} \times s_i^{\alpha} \quad (1)$$

where $s_i = \sum_i \omega_{ij}$, the tuning parameter, α, plays a crucial role in determining the emphasis on node degree or strength. When α falls between 0 and 1, greater significance is attributed to the degree, while a value above 1 prioritizes node strength. Despite its importance, determining the exact α value can be challenging. Wei et al.(30) presented a technique for determining the most suitable value of α. Yustiawan et al.[31] utilized the centrality metric mentioned to identify influential nodes within online social networks. This metric is adaptable to directed networks with weights, incorporating both the node's in-degree and out-degree to determine in-degree and out-degree centrality, respectively. The Degree Centrality (DC) of node i, represented as CD(i), is determined using the following definition:

$$C_D(i) = \sum_j^N a_{ij} w_{ij} \quad (2)$$

where a_{ij} denotes the link or relationship between node i and j and w_{ij} denotes the connection weight.

2.2.2 H-index Centrality(HC):

Lü et al. [32] presented the operator H, specifically designed to act on a restricted set of real numbers $(x_1, x_2, \ldots, x_n)$, producing an integer y. In this context, y signifies the maximum value such that there exist at least y items in $(x_1, x_2, \ldots, x_n)$, with each item being not less than y. Subsequently, the H-index of node i is computed as follows: $h_i = H(k_a, k_b, \ldots, k_c)$ here, k_i represents the node i's degree. and $k_a, k_b, \ldots, k_c$ represents the levels of connectivity among the adjacent nodes.

We set the zero-order H-index, $h_i(0)$, as k_i for a node i. Extending this concept, an n-order H-index (where n > 0) is defined iteratively using the formula:

$$h_i(n) = H(h_a(n-1), h_b(n-1), \ldots, h_c(n-1)) \quad (3)$$

Here, the first-order H-index value is treated as the ultimate H-index value, denoted as $h_i(1) = h_i$.

2.2.3 Betweenness Centrality(BC):

Betweenness centrality (BC) examining the impact of nodes on the information flow along the shortest paths within a network. The betweenness centrality of a specific node i, represented as $C_B(i)$, is formally defined as follows:

$$C_B(i) = \sum_{p,q \neq i} \frac{b_{pq}(i)}{b_{pq}} \quad (4)$$

where b_{pq} is the overall quantity of shortest paths connecting node p to node q, and $b_{pq}(i)$ signifies the count of such shortest paths that pass-through node i.

2.2.4 PageRank(PR):

It evaluates the importance of web pages by analyzing their link structure, operating under the assumption that the impact of a page is gauged by both the number and quality of other pages linking to it When a page receives links from numerous high-quality sources, it is considered to have high quality as well. Zhang et al.[33], define the weighted PageRank as follows:

$$\phi(i) = \gamma \sum_{j \in V} \left(\theta \frac{w_{ji}}{s_j^{(out)}} + (1-\theta) \frac{a_{ji}}{d_j^{(out)}} \right) \phi(j) + \frac{(1-\gamma)\beta_i}{\sum_{i \in V} \beta_i} \quad (5)$$

It excels in networks with a specified direction but is not suitable for application in undirected networks.

2.3 Havrda-Charvat Entropy

To express the entropy of Pythagorean fuzzy sets (PFSs) in a probability-oriented manner, we use Havrda and Charavat's entropy concept ($H^{\Upsilon}(p)$), applying it to a probability mass function represented as $p = \{p_1, p_2, \ldots, p_k\}$,

$$H^{\Upsilon}(p) = \begin{cases} \frac{1}{\Upsilon - 1}\left(1 - \sum_{i=1}^{k} p_i^{\Upsilon}\right), \Upsilon \neq 1 (\Upsilon > 0) \\ -\sum_{i=1}^{k} p_i \log p_i, \Upsilon = 1 \end{cases} \quad (6)$$

Consider a finite set of things denoted as :

$X = \{x_1, x_2, \ldots, x_n\}$ is a finite universe of discourses. Now, for a PFS denoted as P within X, we suggest a specific type of entropy measurement as,

$$E_H^{\Upsilon}(P) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\Upsilon - 1}[1 - ((\lambda^2(x_i))^{\Upsilon} - ((\beta^2(x_i))^{\Upsilon} + ((\Pi^2(x_i))^{\Upsilon}], \Upsilon \neq 1 (\Upsilon > 0) \\ \frac{1}{n}\sum_{i=1}^{n} -(\lambda^2(x_i)\log\lambda^2(x_i) + \beta^2(x_i)\log\beta^2(x_i) + \Pi^2(x_i)\log\Pi^2(x_i)), \Upsilon = 1 \end{cases} \quad (7)$$

where $\lambda^2(x_i)$ is the membership degree, $\beta^2(x_i)$ is the non-membership degree and $\Pi^2(x_i)$ is the indeterminacy of the random variable $x_i \in X$ to the set P.

3. NODE PACK FUZZY INFORMATION CENTRALITY

Here we introduce our proposed method and implemented using an example network. We formulate a complex network model with directed edges with associated weights on them, denoted as G(V, E, W, K). Here, $V = \{v_i\}$ comprises the set of vertices, $E = e_{ij}$,indicates the edges directed from vertex v_i to vertex v_j (i.e $e_{ij} = \{(v_i, v_j)\}$), $W = w_{ij}$ denotes the corresponding weights, and $K = \{(k_{in}(v_i), k_{out}(v_i), k(v_i))\}$ represents the three values associated with each node: in-degree ($k_{in}(v_i)$), out-degree ($k_{out}(v_i)$), and total degree ($k(v_i)$). The calculation method for node degrees in directed networks with weights can be formulated as ,

$$k_{in}(v_i) = \sum_{j=1}^{N} e_{ji} \quad i = 1, 2, \cdots, N$$
$$k_{out}(v_i) = \sum_{j=1}^{N} e_{ij} \quad i = 1, 2, \cdots, N$$
$$k(v_i) = k_{in}(v_i) + k_{out}(v_i) \quad (8)$$

For a given node v_i, $k_{in}(v_i)$ denote the count of incoming edges, representing nodes directed towards v_i, while $k_{out}(v_i)$ indicates the count of outgoing edges, representing nodes directed from

Journal of Theoretical and Applied Information Technology
15th January 2024. Vol.101. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

v_i. It is assumed in this paper that the network is free of loops and multiple edges.

3.1 The Most Direct Path Connecting any Two Vertices within the Network

We can determine the shortest distance between any two nodes in weighted networks represented as ω_{ij}, by applying the Dijkstra algorithm. Consequently, the greatest minimum distance from node i to any other nodes in the network can be defined as follows:

$$L_i = \max_{j \in N, j \neq i} (\omega_{ij}) \tag{9}$$

Here, L_i represents the maximum shortest distance and varies for each node, indicating the locality scale around that specific node.

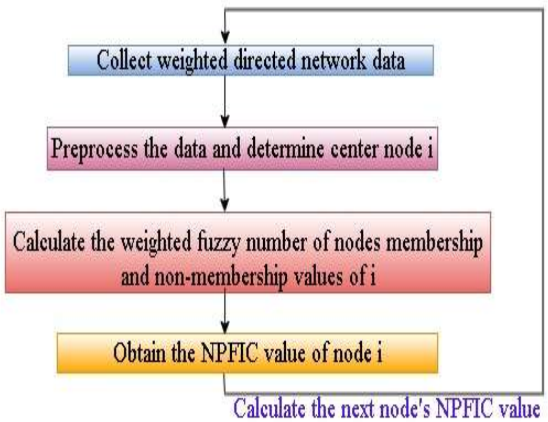

Figure 1: Methodology

3.2 Proposed Method

The significance of a node is evaluated by NPFIC, by taking into account two factors: inbound importance and outbound importance. In this context, the inbound importance of node i is represented by $NPFIC_{in}(i)$, and the outbound importance is denoted by $NPFIC_{out}(i)$. We introduce the definition of NPFIC as outlined below:

Definition : Node Pack Fuzzy Information Centrality

$$NPFIC_{in}(i) = \frac{1}{d^2}\sum_{d=1}^{D} -\Big(\mu^2_{A_{in}}(d)\log\mu^2_{A_{in}}(d) + \vartheta^2_{A_{in}}(d)\log\vartheta^2_{A_{in}}(d) + \pi^2_{A_{in}}(d)\log\pi^2_{A_{in}}(d)\Big) \tag{10}$$

$$NPFIC_{out}(i) = \frac{1}{d^2}\sum_{d=1}^{D} -\Big(\mu^2_{A_{out}}(d)\log\mu^2_{A_{out}}(d) + \vartheta^2_{A_{out}}(d)\log\vartheta^2_{A_{out}}(d) + \pi^2_{A_{out}}(d)\log\pi^2_{A_{out}}(d)\Big) \tag{11}$$

Where the distance 'd' is measured from the central node 'i' and varies within the range of one to the pack size 'D'. NPFIC uses d^2 to account for the time delay in information transmission within a complex network. This approach more accurately models the distribution of communication propagation within complex networks. $\mu^2_{A_{in}}(d)$ and $\vartheta^2_{A_{in}}(d)$ are the membership and the non-membership of neighboring nodes in the in-pack is determined based on the minimal distances to the central node i, which are denoted as d. $\mu^2_{A_{out}}(d)$ and $\vartheta^2_{A_{out}}(d)$ are the membership and the non-membership of neighboring nodes in the out-pack is determined based on the minimal distances from the central node i, which are denoted as d.

The following outlines the comprehensiveness of NPFIC:

Step 1: Acquire the directed weighted network dataset.

Step 2: Load the dataset , preprocess it as needed. Obtain the pack size D by (12). The pack size definition clarifies that NPFIC revolves around the quasi-local information of a node.

$$D = \left\lceil \frac{L_i}{2} \right\rceil \tag{12}$$

where L_i is defined from the eq. (9).

Step 3: Determine the weighted degree of membership of nodes within a fuzzy set, considering the number of neighbor nodes within a specific layer. NPFIC emphasizes the count of directed edges connecting all neighboring nodes in that layer, along with their respective weights. The membership degrees of nodes as weighted fuzzy numbers are provided by,

$$\mu^2_{A_{in}}(d) = \sum_{j=1}^{n_i(d)} \sum_{k=1}^{k_{in}(j)} w_{kj} X(d) \tag{13}$$

$$\mu^2_{A_{out}}(d) = \sum_{j=1}^{n_i(d)} \sum_{k=1}^{k_{out}(j)} w_{jk} X(d) \tag{14}$$

where $k_{in}(j)$, is the quantity of nodes that are directed to node j within the distance d, $n_i(d)$ represents the count of nodes for which the shortest distances from node i equal d, determined through Boolean logic in LID[13] but which is not a reasonable method when dealing with fuzzy data.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

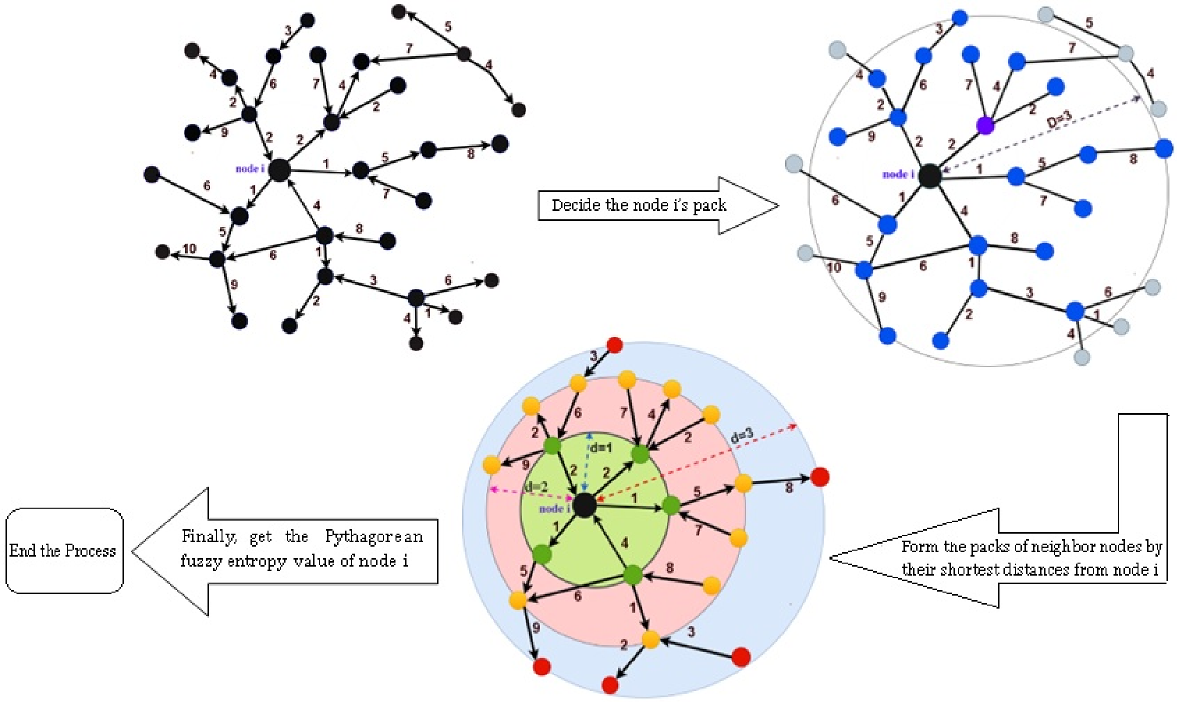

Figure 2: Visual representation of NPFIC's process .

In our method $n_i(d)$ is calculated by using breadth first search algorithm with some improved function. $k_{out}(j)$ is the quantity of nodes emanating from node j within the distance d. Every node is distinguished by its exclusive shortest distances from the central node and acquires diverse weights through fuzzy sets. The assignment of weights to nodes by fuzzy sets is computed based on the following equation:

$$X(d) = exp^{-\frac{d^2}{D^2}} \quad (15)$$

Here, X(d) allows us to assess the contributions of nodes ranked according to their minimal distances from the central node through FSs. Nodes closer to the center have X(d) values closer to one, indicating a higher contribution to the centrality of the node's influence.

Step 4: Calculate non-membership and indeterminacy values by analyzing edge weights, and variability. The non-membership degree of each node in a distance d can be calculated by the inverse of edge weights.

$$\vartheta^2_{A_{in}}(d) = \sum_{j=1}^{n_i(d)} \sum_{k=1}^{k_{in}(j)} \left(1 - \frac{w_{kj}}{L_{i_{in}}}\right) \quad (16)$$

$$\vartheta^2_{A_{out}}(d) = \sum_{j=1}^{n_i(d)} \sum_{k=1}^{k_{out}(j)} \left(1 - \frac{w_{jk}}{L_{i_{out}}}\right) \quad (17)$$

Indeterminacy value can be determined based on the imprecision associated with edge weights. It can be a fixed value or calculated based on the standard deviation of edge weights within a certain neighborhood.

Step5: Calculate the Pythagorean fuzzy entropy of a node i by (10) and (11).

Step6: Define fuzzy membership function by specifying the membership ranges to convert the fuzzy entropy into a fuzzy membership value to node i.

Step7: Rank node i's centrality by organizing nodes according to their fuzzy membership values.

3.3 Example Network Explanation

To demonstrate the functionality of the suggested approach, consider the simple network depicted in figure 3. The diagram illustrates the process of determining the NPFIC value for node i. In the pack of node i, nodes at different shortest distances contain diverse amounts of information. Consequently, the information uncertainty within node i's pack is influenced by the formation is determined by nodes having varied shortest distances from node i. While the NPFIC value for node i is specifically calculated here, it's important to note that in real-time applications, calculating centrality values for each node involves treating each node as a center node.

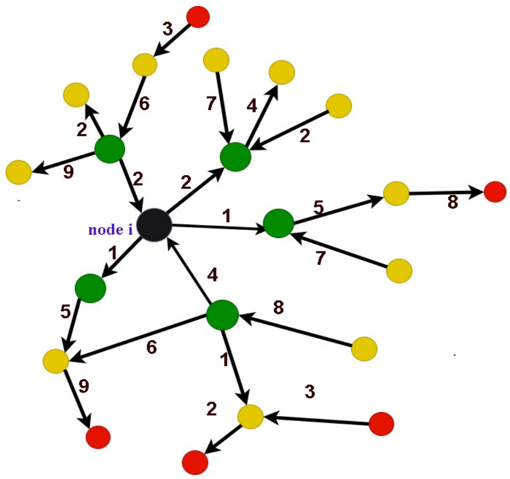

Figure 3: Example Directed Weighted Network

The implementation process is illustrated as follows:

Step 1: Assume that we want to find the centrality of node i as the center node using Pythagorean fuzzy sets with incorporated edge weights and determine its size D in this example suppose 3.

Step 2: NPFIC examines the count of neighboring nodes within a specific layer, emphasizing the directed edges connected to all those neighbors and their respective weights. Compute the weighted distances from node i to all remaining nodes within a 3-hop range, utilizing normalized edge weights.

Step 3: Determine the membership values of nodes $\mu^2_{A_{in}}(d)$ and $\mu^2_{A_{out}}(d)$. Compute the weighted fuzzy values to represent nodes both in in-pack and out-pack using the following calculation,

$$\mu^2_{A_{in}}(1) = \sum_{j=1}^{2}\sum_{k=1}^{2} w_{kj}X(1)$$
$= 0.2 * 0.8948 + 0.4 * 0.8948$
$= 0.5369$

$$\mu^2_{A_{in}}(2) = \sum_{j=1}^{2}\sum_{k=1}^{2} w_{kj}X(2)$$
$= 0.6 * 0.6412 + 0.8 * 0.6412$
$= 0.8976$

$$\mu^2_{A_{in}}(3) = w_{kj}X(3)$$
$= 0.3 * 0.3679$
$= 0.1103$

where $k_{in}(j)$, represents the count of nodes directed towards the central node i, from figure 4(b) and $n_i(d)$ represents the count of neighboring nodes in each layer of the in-pack associated with node i. and w_{kj} is the weight value on each edge in a short distance of d to node i.

$$\mu^2_{A_{out}}(1) = \sum_{j=1}^{3}\sum_{k=1}^{3} w_{jk}X(1)$$
$= 0.1*0.8948+0.1*0.8948+0.2*0.8948$
$= 0.0894 + 0.0894 + 0.1789$
$= 0.3577$

$$\mu^2_{A_{out}}(2) = \sum_{j=1}^{3}\sum_{k=1}^{3} w_{jk}X(2)$$
$= 0.5*0.6412 + 0.4*0.6412 + 0.5*0.6412$
$= 0.3206+0.2564+0.3206$
$= 0.8976$

$$\mu^2_{A_{out}}(3) = \sum_{j=1}^{2}\sum_{k=1}^{2} w_{jk}X(3)$$
$= 0.8*0.3679+0.9*0.3679$
$= 0.2943 + 0.3311$
$= 0.6254$

where $k_{out}(j)$ represents the count of nodes directed away from the central node i, from figure 4(c) and $n_i(d)$ represents the number of adjacent nodes in each layer connected to node i through its outgoing connections and w_{jk} is the weight value on each edge in a short distance of d from node i.

Step 4: Determine non-membership (ϑ^2_A) based on the inverse of edge weights.

$$\vartheta^2_{A_{in}}(1) = \sum_{j=1}^{2}\sum_{k=1}^{2} 1 - \frac{w_{kj}}{L_{i_{in}}}$$
$= (1-\frac{0.2}{0.4})+(1-\frac{0.4}{0.4})$
$= (1-0.5) +(1-1)$
$= 0.5$

$$\vartheta^2_{A_{in}}(2) = \sum_{j=1}^{2}\sum_{k=1}^{2} 1 - \frac{w_{kj}}{L_{i_{in}}}$$
$= (1-\frac{0.6}{1.2})+(1-\frac{0.8}{1.2})$
$= (1 - 0.5)+(1 - 0.6667)$
$= 0.5 + 0.3333$
$= 0.8333$

$$\vartheta^2_{A_{in}}(3) = \left(1 - \frac{0.3}{1.1}\right)$$
$= 1- 0.2727$
$= 0. 7273$

similarly, $\vartheta^2_{A_{out}}(1) = 0.7500$, $\vartheta^2_{A_{out}}(2) = 0.6667$, $\vartheta^2_{A_{out}}(3) = 0.3667$ are also calculated.

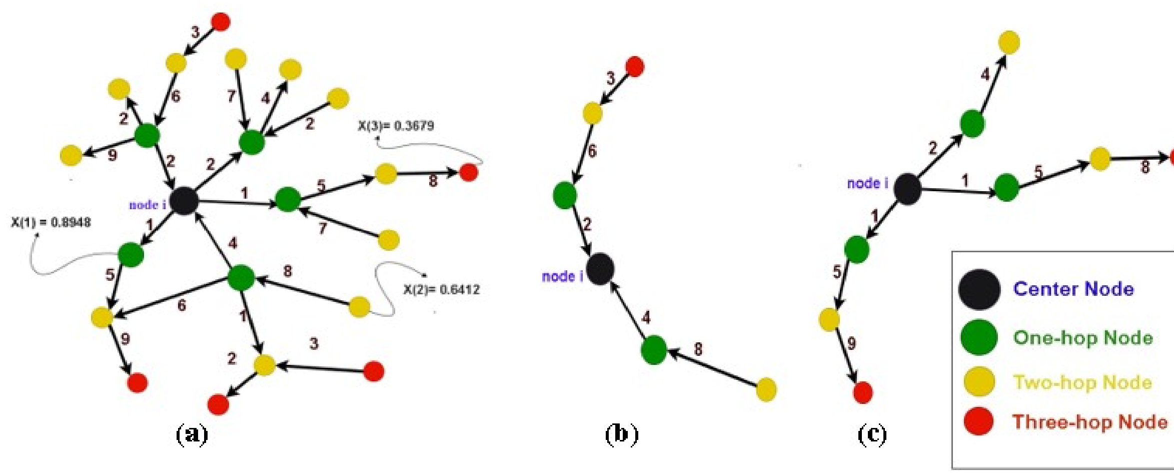

Figure 4: (a) An illustration of a directed weighted pack of Node i. (b) in-pack of Node i. (c) out-pack of Node i.

To determine indeterminacy (π_A^2) , we will use a fixed value in both in-pack and out-pack , let's say π_A^2=0.1, which represents a moderate level of imprecision.

Step 5: Obtain $NPFIC_{in}(i)$ = 0.4813 and $NPFIC_{out}(i)$ = 0.4517 by (10) and (11).

Step 6: Define fuzzy membership function by specifying the membership ranges to convert the fuzzy entropy values to a centrality value of node i. Assume that if the obtained NPFIC value is less than 0.50, it reflects the lower entropy indicates greater importance of a node ,or else if the value less than 0.75 then it reflects moderate centrality, otherwise node has high ambiguity or uncertainty.

Step 7: Finally, determine the Rank Centrality of node i by arranging the centrality scores for each node.

4. RESULTS AND ANALYSIS

We tested the viability and efficiency of NPFIC through experiments on a directed weighted real-world complex network. We compared it against four centrality measures: DC, BC, H-Index, and PageRank.

4.1 Dataset

The effectiveness of NPFIC is verified through its application to a real-world complex network. Soc_bitcoin Network, a weighted directed network comprised of individuals engaging in bitcoin transactions on the social media platform called Bitcoin Alpha. Due to the anonymity of bitcoin users, it is essential to keep a record of user reputation to mitigate transactions with potentially fraudulent and risky individuals. Members on Bitcoin Alpha assess each other by providing ratings that range from total distrust to total trust. In table 1, it presents the details of this network.

Table 1: Fundamental Details About the Dataset

Network	|N|	|V|	$\langle d\rangle$	d_{ma}	$\langle\omega\rangle$	ω_{max}
Soc_bitcoin	3783	24,186	10.345	84	2.567	6

Note: |N| and |V| denote the number of nodes and edges in the dataset. The symbols $\langle d\rangle$ and $\langle d_{max}\rangle$ signify the mean and maximum degree, respectively. Additionally, $\langle\omega\rangle$ and $\langle\omega_{max}\rangle$ represent the mean and maximum distances that are shortest.

4.2 Nodes In Top-10

Here we utilized NPFIC along with four alternative centrality metrics to locate the top ten nodes with the most significant influence in the Soc_bitcoin network. The outcomes are presented in table 2. If a node is colored, it signifies that it ranks within the top ten crucial nodes as identified by NPFIC. Furthermore, a node is underscored when it emerges not only within the top 10 rankings as per NPFIC but also maintains an identical position on that list. It's important to note that variations in fundamental concepts and focal points across various centrality metrics can result in diverse outcomes. In table 2, an analysis of the Soc_bitcoin network indicates that the leading influential nodes identified by NPFIC, H-INDEX, and BC align. PR and NPFIC share identical top-2,

Journal of Theoretical and Applied Information Technology
15th January 2024. Vol.101. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

top-6, and top-7 nodes. There are six nodes consistently ranked in the top 10 between BC and NPFIC, five between DC and NPFIC, and three nodes concurrently ranked in BC, H-INDEX, and NPFIC.

Table 2: Identified Top ten Nodes by Five Centrality Measures.

Rank	Soc_bitcoin				
	DC	BC	PR	H-INDEX	NPFIC
1	41	62	89	62	62
2	62	121	41	73	41
3	51	51	73	51	51
4	63	63	62	63	63
5	89	89	51	58	89
6	73	106	106	110	106
7	121	19	121	115	121
8	16	58	110	41	58
9	114	41	120	86	73
10	110	104	132	28	110

4.3 Assessing the Dissemination Capability through the SI Model

In our study, we employ the susceptible-infected (SI) model[35] to evaluate the effectiveness of various centrality metrics.

4.3.1 SI Model:

The model known as susceptible-infected-recovered(SIR) is a widely used framework for simulating disease transmission in networks. It categorizes individuals into susceptible, infected, and recovered classes. During any specific time interval , the likelihood of a susceptible individual getting affected by neighboring nodes that are also infected is calculated by $c-\left(\frac{1}{2}\right)^{\rho}$, while the probability of the infected recovering is denoted as λ. A relatively small c value ensures that nodes on an individual level contribute to a significant role in the speed and scale of infection. The total population is represented by the sum of susceptible individuals (S(t)), infected individuals (I(t)), and recovered individuals (R(t)). Within the framework of the SI model, which is a modified version of the SIR model, the recovery parameter λ is set to 0, signifying the absence of recovered individuals. In our study, we utilize the SI model to assess the impact of top 10 nodes identified using various centrality measures. These nodes are identified as the network's initially infected nodes, and their impact is observed in the cumulative count of infected nodes across each time interval.

4.3.2 Result Analysis:

In this study, we assigned a value of 2 to ρ to ensure a small spreading infection ability, emphasizing the impact on the top-10 nodes. By conducting 120 independent experiments to mitigate randomness, we enhance result credibility. As depicted in figure 5, initially infected nodes are only ten, aligning with the experiment's context. In the early stages of infection, the overall number of infected nodes undergoes rapid growth, indicating the extensive connectivity between initially infected nodes and numerous nodes that have not yet been affected. As the infection spreads, the growth rate gradually slows down, consistent with the pattern of disease transmission. This trend validates the experiment's rationality and authenticity. In figure 5 illustrates the results for the Soc_bitcoin network, showing that NPFIC consistently outperforms other centrality measures across all time steps, indicating a larger number of infected nodes compared to alternative methods.

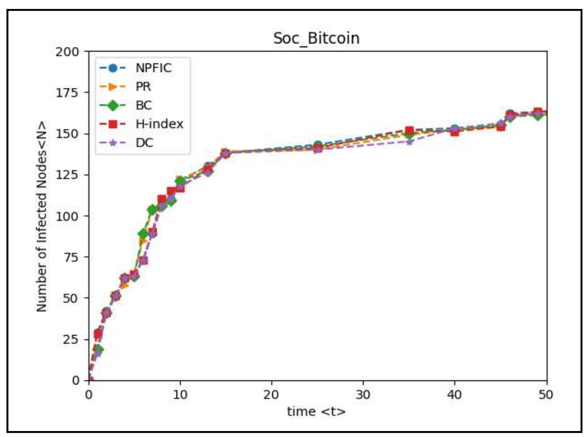

Figure 5: Determine the Average Number of Infected Nodes.

4.4 Evaluate the Network Quality by Eliminating Top Nodes

Evaluate the network's quality by selectively removing crucial nodes, as the connectivity serves as an indicator of the network's overall performance. The configuration of a network is characterized by the following formula:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$Q_n = \frac{\sum_{i}^{N-1}\sum_{j=i+1}^{N} \omega_{ij}}{N(N-1)/2} \quad (18)$$

Let N denote the quantity of nodes within the network, and ω_{ij} denotes the smallest distance between node i and node j. Initially, we eliminated T×10% of elements within the network based on their importance ranking. Subsequently, we update the network and assess its quality using equation (18). If the resulting Q_n value is lower, it indicates a deterioration in the network's connectivity, highlighting the significance of the removed nodes. The nodes identified by NPFIC carry greater influence within the Soc_Bitcoin network compared to those identified through alternative metrics. Due to the NPFIC, the Soc_Bitcoin network exhibits enhanced performance when the T is below 0.50 and above 1.75. Figure 6 illustrates the findings from this experiment.

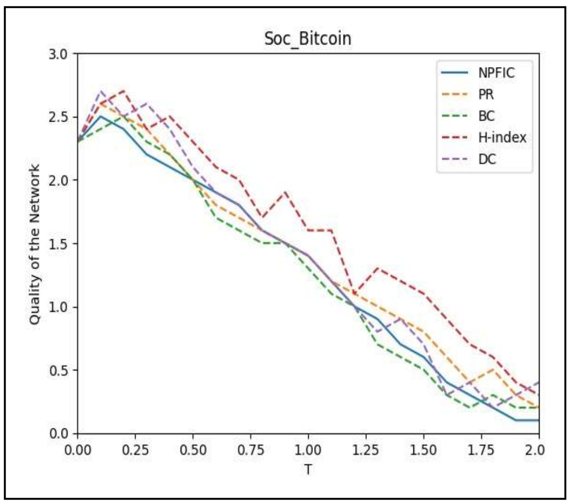

Figure 6: The quality of the network correlates with T.

4.5 Application of NPFIC in E-Commerce Business

This article presents a novel centrality measurement approach employing Pythagorean Fuzzy Sets. In general, the model considers factors such as membership, non-membership, and indeterminacy value of a node, based on these values influence or importance of a node will be calculated. The e-commerce business heavily relies on central nodes, and while existing methods focus on a single parameter like node connectivity, our innovative approach considers two crucial factors: node connectivity (from directions) and self-weight (from activities). In the realm of e-commerce and spreading news, self-weight emerges as a paramount parameter. Consequently, our model proves to be more effective for e-commerce businesses. In figure 7, we depict a compact network representing online e-commerce platforms such as Croma, Amazon, Flipkart, Meesho, and others. In this network, every seller and customer are treated as a vertex, connected by edges representing purchases.

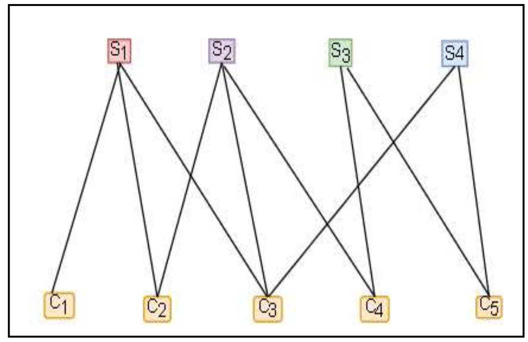

Figure 7: A small online e-commerce Network.

Seller vertex values are determined by positive and negative reviews, as well as indeterminacy from non-reviews. Customer vertex values rely on monthly product purchases, order cancellations, and inactivity. Edge membership values are assigned based on matches. Our model NPFIC,promises enhanced outcomes for online e-commerce ventures.

5. CONCLUSION

In this article we present an innovative method known as NPFIC to detect influential nodes in complex networks, by focusing on the internal arrangement of a node's pack to determine its global significance. Unlike existing approaches, NPFIC considers that the influence of neighboring nodes on a central node's importance is closely tied to their degree of membership and non-membership, so it employs fuzzy sets to distribute different weights to neighbor nodes within a node's pack. NPFIC uses Havrda-Charvat entropy, the probability-type of entropy for PFSs to better suit real-world directed weighted network characteristics. NPFIC assesses node importance by examining the information contained within the node's pack defined by the network structure which are all calculated through the Havrda-Charvat entropy. In computing information, NPFIC considers the inner structure of node's pack, providing a more accurate and refined approach to evaluate node importance globally in complex networks. The effectiveness and superiority of NPFIC are confirmed through several experimental validations done on complex networks, where it is

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

contrasted with other four centrality metrics. The nodes ranked in the top 10 based on NPFIC demonstrate their importance through a wider range of influence compared to four other methods. In conclusion, NPFIC not only helps in identifying influential nodes within directed weighted networks but also effectively integrates principles from information theory with the field of network science. Expanding the utilization of NPFIC to Neutrosophic graphs within social networks aims to identify influential actors and assess the uncertainty associated with the contributions of neighboring nodes. This endeavor serves to inspire and guide future research in this domain.

REFERENCES:

[1] Newman MEJ,"The structure and function of complex networks", SIAM Rev 2003;45(2) pp:167–256. doi:10.1137/s003614450342480.

[2] L. Zadeth, "Fuzzy sets," Inf. Control, vol. 8, 1965, pp. 338–353,.

[3] Z. Li, X. Liu, J. Dai, J. Chen, and H. Fujita, "Measures of uncertainty based on Gaussian kernel for a fully fuzzy information system" ,Knowl.-Based Syst., vol. 196, 2020, Art. no. 105791.

[4] L. Wang and H. Garg, "Algorithm for multiple attribute decision-making with interactive archimedean norm operations under pythagorean fuzzy uncertainty" , Int. J. Comput. Intell. Syst., vol. 14, no. 1,2020, pp. 503–527.

[5] R. C. de Moura, G. B. Schneider, L. de Souza Oliveira, M. L. Pilla, A. C. Yamin, and R. H. S. Reiser, "f-hybridmem: A fuzzy-based approach for decision support in hybrid memory management" , in Proc. IEEE Int. Conf. Fuzzy Syst., 2020, pp. 1–8.

[6] X. Hu, W. Pedrycz, and X. Wang, "Granular fuzzy rule-based models: A study in a comprehensive evaluation and construction of fuzzy model," , IEEE Trans. Fuzzy Syst., vol. 25, no. 5, Oct. 2017, pp. 1342–1355.

[7] X. Peng and H. Garg, "Multiparametric similarity measures on pythagorean fuzzy sets with applications to pattern recognition" , Appl. Intell., vol. 49, no. 12, 2019, pp. 4058–4096.

[8] R. R. Yager, "Pythagorean membership grades in multicriterion decision making" , IEEE Transactions on Fuzzy Systems, vol. 22, no. 4, 2014, pp. 958–965.

[9] X. L. Zhang and Z. S. Xu, "Extension of TOPSIS to multiple criteria decision making with pythagorean fuzzy sets" ,International Journal of Intelligent Systems, vol. 29, no. 12, 2014, pp. 1061–1078.

[10] W. Xue, Z. Xu, X. Zhang, and X. Tian, "Pythagorean fuzzy LINMAP method based on the entropy theory for railway project investment decision making", International Journal of Intelligent Systems, vol. 33, no. 1, 2018, pp. 93–125.

[11] T. Wen and K. H. Cheong, "The fractal dimension of complex networks: A review" , Inf. Fusion, vol. 73, 2021, pp. 87–102.

[12] A. Zareie, A. Sheikhahmadi, and A. Fatemi, "Influential nodes ranking in complex networks: An entropy-based approach" , Chaos, Solitons Fractals, vol. 104,2017, pp. 485–494.

[13] T. Wen and Y. Deng, "Identification of influencers in complex networks by local information dimensionality", Inf. Sci., vol. 512, 2020, pp.549–562.

[14] M. E. Newman, "The structure and function of complex networks" , SIAM Rev., vol. 45, no. 2,2003, pp. 167–256.

[15] M. E. Newman, "A measure of betweenness centrality based on random walks" ,Social Netw., vol. 27, no. 1,2005, pp. 39–54.

[16] L. C. Freeman, "Centrality in social networks conceptual clarification" , Social Netw., vol. 1, no. 3, 1978, pp. 215–239.

[17] Bonacich P (1987) "Power and centrality: a family of measures" ,Am J Sociol 92(5):1170–1182.

[18] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine" , Comput. Netw. ISDN Syst., vol. 30, no. 1–7, 1998, pp. 107–117.

[19] Wang P, Lu J, Yu X (2014) " Identification of important nodes in directed biological networks: a network motif approach" , PLoS ONE 9(8):e106132.

[20] Sheng J, Dai J,Wang B, Duan G, Long J, Zhang J, Guan K, Hu S, Chen L, Guan W (2020) "Identifying influential nodes in complex networks based on global and local structure" , Physica A 541:123262.

[21] Wang X, Slamu W, Guo W, Wang S, Ren Y (2022) "A novel semi local measure of identifying influential nodes in complex. Networks", Chaos Solitons Fractals 158:112037.

[22] Panfeng L, Li L, Shiyu F, Yukai Y (2021) “Identifying influential nodes in social networks: a voting approach”, Chaos Solitons Fractals 52:111309.

[23] Antonios Garas et al ,"A k-shell decomposition method for weighted networks" , New Journal of Physics 14 (2012) 083030.

[24] R. R. Yager and A. M. Abbasov, “Pythagorean membership grades, complex numbers, and decision making” , International Journal of Intelligent Systems, vol. 28, no. 5, 2013, pp. 436–452.

[25] L. Lü, T. Zhou, Q.-M. Zhang, H.E. Stanley, “The H-index of a network node and its relation to degree and coreness”, Nature Commun. (1) (2016) 1–7, http://dx.doi.org/10.1038/ncomms10168.

[26] Qiao, T.; Shan, W.; Zhou, C. “How to Identify the Most Powerful Node in Complex Networks? A Novel Entropy Centrality Approach”, *Entropy* 2017, *19*,614. https://doi.org/10.3390/e19110614

[27] G. Sabidussi, “The centrality index of a graph”, Psychometrika 31 (4) (1966) 581–603, http://dx.doi.org/10.1007/BF02289527.

[28] M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, H.A.Makse, “Identification of influential spreaders in complex networks” ,Nat. Phys. 6 (11) (2010) 888–893.

[29] Tore Opsahl, Filip Agneessens, and John Skvoretz, “Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks” ,32(3),2010,pp:245–251.

[30] Daijun Wei, Ya Li, Yajuan Zhang, and Yong Deng. “Degree centrality based on the weighted network” , In Control and Decision Conference (CCDC), 2012 24th Chinese, IEEE, 2012, pages 3976–3979.

[31] Yoga Yustiawan, Warih Maharani, and Alfian A. Gozali. “Degree Centrality for Social Network with Opsahl Method” , Procedia Computer Science, 2015,59:419–426.

[32] Lü, L., Zhou, T., Zhang, Q. M., & Stanley, H. E. (2016). “The H-index of a network node and its relation to degree and coreness” , *Nature communications*, *7*(1), 10168.

Stat. Mechanics Appl., vol. 390, no. 12, 2011,pp. 2408–2413.

[33] P.Zhang, Tiandong Wang and Jun Yan ‘‘PageRank centrality and algorithms for weighted,directed networks with applications to World Input-Output Table”, [Phys. Soc-ph.], Stat. Mech. Appl., vol. 444, May. 2021, pp. 73–85.

[34] J. Havrda and F. Charv´at, “Quantication method of classication processes. Concept of structural α-entropy”, Kyberneti k a ,vol. 3, 1967, pp. 30–35.

[35] M. Yang, G. Chen, and X. Fu,“A modified sis model with an infective medium on complex networks and its global stability” ,Physica A: