IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 January 2024, accepted 31 January 2024, date of publication 7 February 2024, date of current version 16 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3363635

RESEARCH ARTICLE

An Enhanced Gravity Model for Determining Crucial Nodes in Social Networks Based on Degree K-Shell Eigenvector Index

HARDEEP SINGH[1] AND HARDEEP SINGH[2], (Member, IEEE)
[1]Department of Computer Engineering and Technology, Guru Nanak Dev University, Amritsar 143005, India
[2]Department of Computer Science, Guru Nanak Dev University, Amritsar 143005, India

Corresponding author: Hardeep Singh (hardeep.cet@gndu.ac.in)

ABSTRACT In the realm of network science, determining crucial nodes within a social network is an ongoing concern. As a result, it garners a lot of attention, and various centrality measures for the identification of crucial nodes have been proposed thus far. Degree and k-shell decomposition are the classic centrality measures that rely on neighboring nodes. However, degree, k-shell, and combination of degree and k-shell measures assign the identical value to the vast count of nodes, which creates a problem in distinguishing these nodes. Therefore, in this paper, for the purpose of solving the above problem, we propose an index based on three different components: degree, improved k-shell measure, and eigenvector centrality called the degree k-shell eigenvector (DKE) index. In addition, we propose an enhanced gravity model called the DKE-based gravity model (DKEGM) on the basis of universal gravity law and the proposed index for determining crucial nodes in social networks. The proposed gravity model incorporates different aspects of nodes, which include count of neighbors, location of nodes, influence of neighbors, and path information between the nodes. Numerous experiments are executed on eight real networks using the SIR model, Kendall tau, ranking monotonicity, and distinct metric to examine the effectiveness of the DKEGM with respect to the other measures. The empirical outcomes show the effectiveness of the DKEGM in terms of accuracy, distinguishing ability, and efficiency.

INDEX TERMS Centrality measures, crucial nodes, degree k-shell eigenvector index, gravity model, social networks.

I. INTRODUCTION

Social networks are representations of real social systems in which individuals are denoted as nodes and the links between individuals are denoted as edges. Millions of individuals have cell phones, and they use different social media apps like WhatsApp or Facebook to communicate and exchange information with family or friends; LinkedIn to look for jobs; Twitter to report news; and Instagram to make and share video reels. As a result, it serves as a platform for advertising products, sharing news, and so on. To execute the above-mentioned works, finding the crucial nodes is an essential task because, in social networks, a node with higher spreading capability is considered crucial, which means it is able to disseminate information to a large number of users [1]. Most studies only pay attention to a lower count of crucial nodes than the entire network because of the vast amount of data [2]. Crucial nodes have a stronger influence on the network's structure as compared to the rest of the nodes [3]. So, finding the most crucial nodes helps in effectively analyzing the whole network. Crucial nodes can be mined for both theoretical and practical purposes. For instance, to prevent the transmission of viruses and rumors [4], [5], speed the distribution of effective information throughout the network [6],

The associate editor coordinating the review of this manuscript and approving it for publication was Feiqi Deng.

etc. In a network, centrality denotes the node that is most crucial [7]. So, figuring out centrality in a social network is very essential.

Numerous centrality measures have been put forth because of the huge increase in the social networks, including degree [8], eigenvector [8], closeness [9], betweenness [10], k-shell decomposition [11], gravity index [12], etc. These centrality measures are categorized as local and global [13]. Local information about the node is used to generate local measures of centrality. For example, degree measure of centrality, described as the count of closest neighbors, is fast to compute but has low accuracy. On the other hand, centrality measures that necessitate the entire structure of a network are termed global centrality measures because they cannot be computed without the use of global information. For example, centralities such as closeness, betweenness, eigenvector, etc. Closeness centrality is more accurate than degree centrality since it takes into account direct as well as indirect links between nodes, but it is not appropriate for networks involving disconnected subgraphs. Betweenness and closeness centralities are not appropriate for networks of enormous size because of their high time complexity. Both the count and influence of a node's neighbors compute a node's influence as per the eigenvector measure of centrality, but it performs well only if the network is connected. Later, another global measure of centrality called the k-shell method was introduced to evaluate a node's influence according to its location. Using an iterative process, this method divides each and every node into k-shells. The highest value of the k-shell represents the most crucial node and is situated near the network's core. Core nodes are more crucial as compared to non-core nodes. It is unable to differentiate between crucial nodes within the common core level.

Although the above-mentioned centrality measures have some advantages, due to their limitations, it is hard to determine the crucial nodes in social networks correctly and effectively. As a result, numerous centralities have been introduced in recent years, but the centrality measures depend on the universal gravity formula have shown effective results [14]. The gravity index centrality [12], local gravity model [15], and DK-based gravity model [14] are typical examples. According to the gravity index centrality, a node's k-shell value represents its mass, while the shortest path between two nodes represents their distance. The local gravity model uses a node's degree value to represent its mass and uses the shortest path between two nodes to represent their distance. The DK-based gravity model depends on the degree k-shell (DK) index. This index combines both degree and improved k-shell measures. The DK-based gravity model uses a node's DK index value to represent its mass and uses the shortest path between two nodes to represent their distance. But all these gravity models have a problem in distinguishing the nodes because degree, k-shell, and the DK index assign identical values to the vast count of nodes. Also, degree and k-shell measures only find the count of nearest neighbors and the location of nodes, respectively, but both do not consider the influence of neighbors. Different neighbors may have a different impact on determining an individual's spreading capability, as noted in [16]. This is due to the heterogeneous structure of social networks, which restricts every node from having the same importance inside the network [17]. In brief, creating an effective measure of centrality that encompasses key aspects of nodes for determining the crucial nodes within a network remains an open problem.

In this study, for the purpose of solving the above problem, we propose the degree k-shell eigenvector (DKE) index to create an effective gravity model. This index consists of three different components of a node: the count of its neighbors (degree), its location (improved k-shell measure), and the influence of its neighbors (eigenvector centrality). We consider these three components together because a node is crucial, not just because it has many neighbors but also because those neighbors are crucial [18]. In addition, a node is said to be in the network's core position if all of its neighbors are also in the core position [3]. This indicates that the importance of a node is based on its own position and also on the position of its neighbors. Therefore, the above-mentioned three different components of the proposed index consider distinct aspects of a node's importance, which results in the high distinguishing ability of the proposed index. Furthermore, we propose an enhanced gravity model called the DKE-based gravity model (DKEGM) on the basis of the universal gravity law and the proposed index for determining crucial nodes in social networks. This model incorporates different aspects of nodes, which include count of neighbors, location of nodes, influence of neighbors, and path information between the nodes. The SIR model, Kendall tau, ranking monotonicity, and distinct metric are utilized as evaluation criteria on eight real networks to determine the efficacy of the DKEGM. Empirical outcomes show that the DKEGM is superior to other centralities that include closeness, degree, eigenvector, betweenness, k-shell, global and local information, gravity index, local gravity model, and the DK-based gravity model.

The main advantages of the proposed model (DKEGM) are as follows:

- **Accuracy:** The DKEGM comprehensively incorporates distinct aspects of a node's importance, which helps to determine the crucial nodes more accurately and effectively.
- **Distinguishing Ability:** Compared with other centralities, the DKEGM has a high ability to distinguish the nodes' influence.
- **Scalability:** The DKEGM has low time complexity as it utilizes two-order neighbors of nodes to determine crucial nodes. Therefore, it is suitable for complex and large social networks.
- **Robustness:** The DKEGM makes our research more robust. If a single measure has some limitations, then a

more balanced assessment of nodes can be obtained by integrating different measures.

The remaining portion of this study is structured as: We summarize the previous research in Section II. We introduce some past and recent proposed measures of centrality in Section III. Furthermore, we present our proposed model in Section IV. In Section V, we focus on the experimentation and analysis of the results. In Section VI, we come to a conclusion and provide some suggestions for the future.

II. RELATED WORK

In the past few years, it has become challenging to determine the crucial nodes. For identifying crucial nodes, numerous centralities have been developed, and each measure has pros and cons. In order to examine communication networks, Bavelas was the first to develop the centrality measure for connected graphs and propose its use [19]. Degree, betweenness, and closeness are three mathematical models of centrality developed by Freeman [10]. The simplest measure with low time complexity and low accuracy is known as degree centrality. It shows how many nodes are directly connected to a specific node. Closeness and betweenness centralities are not efficient for networks of enormous size due to their high time complexity. Reference [8] proposed an eigenvector measure in which a node is considered crucial if it has crucial neighbors. Kitsak et al. [11] suggested the method termed k-shell for the identification of crucial nodes. A node's influence is computed using this method according to its location. Bae and Kim [20] suggested a coreness measure through which the nodes' influence is computed using their neighbors' k-shell values.

To discover the most crucial nodes within a network, [12] presented a gravity index centrality and an extended gravity index centrality based on the principle of gravity law. These gravitational centrality measures have high accuracy, but their computational complexity is also high for networks of large size. Berahmand et al. created a semi-local measure of centrality that is appropriate for networks of large size due to its near-linear time complexity [21]. Also, Berahmand et al. found that degree centrality is preferable over other measures for determining crucial nodes in networks that have a high rich-club due to the utilization of local information and its linear time complexity [22]. Fei et al. proposed a measure called inverse-square law centrality to determine the intensity of a node, which is dependent on the total of a node's attraction to all the other nodes inside the network [23].

Yu et al. [24] developed a measure termed ProfitLeader, which computes the nodes' influence based on their profit capacity. When a node generates more profit for others, it becomes more crucial. Reference [15] proposed a local gravity model that also depends on the universal gravity law. It lowers the time complexity by providing the truncation radius. Dai et al. proposed a centrality called local neighbor contribution, which is simple to compute and appropriate for large networks [25]. For discovering the crucial nodes, a method called local and global influence [26] was created. It makes use of each node's location both locally and globally to evaluate its influence. The global structure model [27], which incorporates the influence of the nodes both locally and globally within a network, was developed in order to identify crucial nodes.

Reference [28] proposed an improved gravity measure of centrality that discovers the crucial nodes according to the k-shell method. Reference [29] developed a local and global measure of centrality that accounts for both local and global structural attributes of a network and is used to discover crucial nodes in complex networks. Reference [14] developed an improved version of the gravity model that depends on the degree k-shell (DK) index. It was suggested to leverage the information both locally and globally using the escape velocity measure of centrality and an extended escape velocity measure of centrality [3], which are based on the principle of escape velocity. Hu et al. created a measure for determining crucial nodes called global and local information [30]. For determining vital nodes, a measure termed hybrid characteristic centrality and its extended version were presented [31]. Different research works on centrality measures over the years are summarized in Table 1, and a list of symbols utilized in this study is shown in Table 2.

III. PRELIMINARIES

In this section, we present some previous and recent proposed measures of centrality. An unweighted and undirected social network $G = (V, E)$, where V and E indicate the set of nodes and edges, respectively. A social network G contains $N = |V|$ nodes and $M = |E|$ edges, and $A = \left(a_{pq}\right)_{N\times N}$ indicates the adjacency matrix of G, where a_{pq} indicates the element of matrix A. The value of a_{pq} is 1 if p and q are linked, otherwise 0. Different measures of centrality are described as below:

Degree centrality (DC) is the classic measure to compute a node's influence by counting only its nearest neighbors [8]. The DC [10] is expressed for node p as:

$$DC\left(p\right) = \frac{\sum_{q=1}^{N} a_{pq}}{N-1} \tag{1}$$

where N indicates the count of the nodes.

Betweenness centrality (BC) tracks "how many times a node along the shortest path acts as a bridge between two other nodes". According to this centrality, a node is considered crucial if there are a large count of shortest paths that go through it [10]. It is expressed for node p as:

$$BC\left(p\right) = \frac{2}{\left(N-1\right)\left(N-2\right)} \sum\nolimits_{u\neq p, u\neq v, p\neq v} \frac{g_{uv}\left(p\right)}{g_{uv}} \tag{2}$$

where $g_{uv}\left(p\right)$ stands for the overall count of shortest paths that go via node p in order to join nodes u and v, and g_{uv} stands for the overall count of shortest paths that join nodes u and v.

Closeness centrality (CC) is "the inverse of a node's farness", and the summation of a node's distances from every other node determines the farness of a node [9]. Therefore,

TABLE 1. Related work for identifying crucial nodes using distinct centralities.

S. No.	Author	Year	Centrality Measures	Topology	Description	Advantages	Limitations/Future Work
1.	Bonacich [8]	1972	Eigenvector centrality	Global	Demonstrates that a node is considered crucial if it has crucial neighbors	Considers both the count and influence of neighbors	Performs well only if the network is connected
2.	Freeman [10]	1978	Degree centrality	Local	Considers only the nearest neighbors of a node	Simplest measure with low time complexity	Low accuracy
3.	Freeman [10]	1978	Closeness centrality	Global	Includes both the direct and indirect links between nodes	Identify nearest nodes	High time complexity and inappropriate for networks involving disconnected subgraphs
4.	Freeman [10]	1978	Betweenness centrality	Global	Tracks "how many times a node along the shortest path acts as a bridge between two other nodes"	Find the nodes that serve as bridges between various network segments	High time complexity and unable to distinguish most boundary nodes
5.	Kitsak *et al.* [11]	2010	k-shell decomposition	Global	Considers the location of a node within the network	Low time complexity	Cannot distinguish between crucial nodes within the common core level
6.	Bae and Kim [20]	2014	Coreness centrality	Global	Uses the k-shell values of a node's neighbors to calculate its influence	Better monotonic ranking and accuracy	Applying the coreness measure to weighted and directed networks still an open problem
7.	Ma *et al.* [12]	2016	Gravity index centrality and its extended version	Global	Depends on the universal gravity law, in which a node's k-shell value represents its mass, while the shortest path between two nodes represents their distance	Better accuracy	High time complexity
8.	Berahmand *et al.* [21]	2018	Semi-local and free-parameter centrality	Semi-local	Finds structural holes by integrating the degree, positive effects of second-level neighbors' clustering coefficient of a node, and negative effects of the clustering coefficient of a node	Linear time complexity	Appropriate only for undirected and unweighted networks
9.	Ibnoulouafi and Haziti [32]	2018	Density centrality	Global	Based on the area density formula, which identifies the crucial nodes by including the degree and the shortest distance between two nodes	Better accuracy	Applying density centrality to directed and weighted networks still an open problem
10.	Wang *et al.* [33]	2018	Improved gravity centrality and its extended version	Global	Depends on the universal gravity law, in which a node's k-shell value represents its mass and the degree value represents the mass of its neighbors	Better accuracy	High time complexity
11.	Fei *et al.* [23]	2018	Inverse-square law centrality	Global	Computes a node's intensity, which depends on the total of a node's attraction to every other node within the network	Better accuracy	Cannot differentiate the nodes with the same intensities
12.	Yu *et al.* [24]	2019	ProfitLeader	Local	Computes a node's influence according to its profit capacity, which means if a node generates more profit for others, it becomes more crucial	Simple to compute with low time complexity	Does not work very well for small networks

TABLE 1. ***(Continued.)*** **Related work for identifying crucial nodes using distinct centralities.**

13.	Li *et al.* [15]	2019	Local gravity model	Semi-local	Depends on the universal gravity law, in which a node's degree value represents its mass, while the shortest path between two nodes represents their distance	Lowers the time complexity by providing the truncation radius R	To determine the additional parameter that is truncation radius R
14.	Dai *et al.* [25]	2019	Local neighbor contribution	Local	Considers the contribution of the nearest and next nearest neighbors of the node together with the node's own contribution to determine a node's influence	Low time complexity	Enhances accuracy up to some extent
15.	Sheng *et al.* [34]	2020	Global and local structure	Global	Takes into account the network's global structure, which includes the node's capacity to exchange information with the rest of the nodes, as well as its local structure, which encompasses the node's own influence	Better accuracy	High time complexity
16.	Li *et al.* [35]	2021	Generalized gravity model	Semi-local	Considers both the local clustering coefficient and degree while evaluating the local information of each node	More precise local information	To find the additional parameter α more accurately is still an open problem
17.	Qiu *et al.* [26]	2021	Local and global influence	Global	Depends on the local influence, which includes both the local clustering coefficient and degree, and the global influence, which includes the k-shell method	Better distinguishing ability and accuracy	Less efficiency
18.	Yang and Xiao [28]	2021	k-shell gravity centrality	Global	Incorporates the global information, such as the node's location and its distance from other nodes, as well as the local information, such as the node's degree	More comprehensive measure	Applying k-shell gravity centrality to weighted networks still an open problem
19.	Shang *et al.* [36]	2021	Effective distance gravity model	Global	Incorporates both static and dynamic information by using the effective distance to determine the crucial nodes	More comprehensive measure	High time complexity
20.	Ullah *et al.* [27]	2021	Global structure model	Global	Incorporates the node's self-influence, which includes the node's k-shell value, as well as the node's global influence, which includes the k-shell value of its neighboring nodes along with the shortest distance between the nodes	Better Accuracy	High time complexity
21.	Ullah *et al.* [29]	2021	Local and global centrality	Global	Considers two factors: the node's degree to determine the local influence, as well as the neighboring nodes' degree and the shortest path connecting them to determine the global influence	Better Accuracy	Appropriate only for undirected and unweighted networks
22.	Li and Huang [14]	2021	DK-based gravity model	Global	Depends on the degree k-shell (DK) index and the universal gravity law, which determines the crucial nodes in complex networks	Better accuracy and distinguishing ability	Use of the DK-based gravity model on weighted networks still an open problem
23.	Ullah *et al.* [3]	2022	Extended escape velocity centrality	Global	Utilizes the escape velocity formula by considering both the local feature, which includes the degree, and the global feature, which includes the location and the shortest distance between the nodes	Better Accuracy	Suitable only for unweighted and undirected networks

TABLE 1. *(Continued.)* **Related work for identifying crucial nodes using distinct centralities.**

24.	Hu *et al.* [30]	2022	Global and local information	Global	Determines the crucial nodes by considering the global information, which includes the location, and the local information, which includes the degree	Better Accuracy	Only considers the information of adjacent nodes
25.	Liu and Zheng [31]	2023	Extended hybrid characteristic centrality	Global	Uses multiple characteristics of nodes, including an extended degree and E-shell decomposition, for determining the influence of nodes	Better accuracy and distinguishing ability	Applicable only to undirected and unweighted networks

the smaller a node's total distance from other nodes, the more crucial it will be. The CC [10] is expressed for node p as:

$$CC\left(p\right)=\frac{N-1}{\sum_{q\neq p}d(p,q)} \tag{3}$$

where $d(p,q)$ stands for the shortest path distance that connects nodes p and q.

Eigenvector centrality (EC) indicates that both the count and importance of a node's neighbors determine its influence [8]. It is expressed for node p as:

$$Ay=\lambda y,EC\left(p\right)=y_p=\frac{1}{\lambda}\sum_{q=1}^{N}a_{pq}y_q \tag{4}$$

where y_p denotes the value of the pth item of the eigenvector y with respect to the greatest eigenvalue λ of the matrix A.

k-shell decomposition (KS) is utilized to compute the influence of a node as per its location [11]. The value of k-shell is allotted to each node during the decomposition process. The most crucial node is the one that has the highest value of the k-shell and that provides the greatest potential to spread information. Despite having the advantage of low time complexity, the ranking outcome of this measure is coarse-grained since it cannot distinguish between crucial nodes within the common core level.

Gravity index centrality (GIC) depends on the universal gravity formula, which identifies the crucial nodes by using information regarding the path and neighborhood of the nodes [12]. It is expressed for node p as:

$$GIC\left(p\right)=\sum_{q\epsilon\psi(p)}\frac{KS\left(p\right)KS\left(q\right)}{d^2\left(p,q\right)} \tag{5}$$

where $KS\left(p\right)$ and $KS\left(q\right)$, respectively, stand for the values of the k-shell for nodes p and q, and $\psi(p)$ is a set of neighboring nodes of node p.

Local gravity model (LGM) also works on the principle of gravity law to evaluate a node's influence by adding up all of a node's interactions with other nodes in the network. It presents the truncation radius to lower the time complexity of large networks [15]. It is expressed for node p as:

$$LGM\left(p\right)=\sum_{d(p,q)\leq R,q\neq p}\frac{D\left(p\right)D\left(q\right)}{d^2\left(p,q\right)} \tag{6}$$

where $D\left(p\right)$ and $D\left(q\right)$, respectively, stand for the degree values of nodes p and q, R stands for the truncation radius, and the degree is expressed for node p as:

$$D\left(p\right)=\sum_{q=1}^{N}a_{pq} \tag{7}$$

R can be estimated as:

$$R\approx\frac{1}{2}<d> \tag{8}$$

where $<d>$ denotes the average shortest distance.

DK-based gravity model (DKGM) depends on the degree k-shell (DK) index [14]. This index combines both degree and improved k-shell measures. The DKGM uses both path-based and neighborhood-based information to determine a node's influence. To enhance the distinguishing ability of the nodes, it presents the stage number when a node is eliminated from the network while performing the k-shell decomposition. It is expressed for node p as:

$$DKGM\left(p\right)=\sum_{d(p,q)\leq R,q\neq p}\frac{DK\left(p\right)DK\left(q\right)}{d^2\left(p,q\right)} \tag{9}$$

where $DK\left(p\right)$ and $DK\left(q\right)$, respectively, stand for the DK index values of nodes p and q, and the DK index is expressed for node p as:

$$DK\left(p\right)=D\left(p\right)+KS^*\left(p\right) \tag{10}$$

where $D\left(p\right)$ indicates the value of degree for node p, $KS^*\left(p\right)$ indicates the value of improved k-shell for node p, and $KS^*\left(p\right)$ can be expressed as:

$$KS^*\left(p\right)=KS\left(p\right)+\frac{R\left(p\right)}{maxS\left(k\right)+1} \tag{11}$$

where $KS\left(p\right)$ indicates the value of k-shell for node p, and during the k-shell decomposition process, the overall count of stages is $S\left(k\right)$ for the k-degree iteration, and node p is eliminated in the $R\left(p\right)$ stage.

Global and local information (GLI) is a technique for determining crucial nodes and considers the influence both globally and locally [30]. It can be expressed for node p as:

$$GLI\left(p\right)=GI\left(p\right)+LI\left(p\right) \tag{12}$$

where $GI\left(p\right)$ and $LI\left(p\right)$ stand for the global influence and local influence of node p, respectively.

TABLE 2. Table of symbols.

Symbol	Definition
G	Undirected and unweighted social network
V	Set of nodes
E	Set of edges
N	Count of nodes
M	Count of edges
D	Degree
$g_{uv}(p)$	Overall count of shortest paths that join nodes u and v via node p
g_{uv}	Overall count of shortest paths that join nodes u and v
$d(p, q)$	Shortest path distance that connects nodes p and q
R	Truncation radius
R^*	Optimal truncation radius
$< d >$	Average shortest distance
ψ	Set of neighboring nodes
$R(p)$	Stage at which node p is eliminated
$S(k)$	Overall count of stages
DK	Index that combines both degree and improved k-shell measures
GI	Global influence
LI	Local influence
I	Personal influence
$A(p)$	Influence of closest nodes of node p
$SUM(p)$	Total aggregate of the influence of the closest nodes of node p
$\max D$	Maximal node degree
$\Pi(q)$	Influence of closest node q
Γ	Set of closest nodes
$Jacc(p, q)$	Jaccard similarity coefficient of node p and node q
$n(p)$	Set of nodes that contain links with node p
DKE	Index that combines degree, improved k-shell, and eigenvector measures
DC	Degree centrality
BC	Betweenness centrality
CC	Closeness centrality
EC	Eigenvector centrality
KS	k-shell measure
KS^*	Improved k-shell measure
GIC	Gravity index centrality
LGM	Local gravity model
$DKGM$	DK-based gravity model
GLI	Global and local information
$DKEGM$	DKE-based gravity model
β	Infection probability
β_{th}	Infection threshold

TABLE 2. ***(Continued.)*** **Table of symbols.**

μ	Recovery probability
$< K >$	Average degree
$< K^2 >$	Second-order average degree
K_{max}	Maximum degree
$< CC >$	Average clustering coefficient
σ	Density
τ	Kendall tau correlation coefficient
n_c	Overall count of concordant pairs
n_d	Overall count of discordant pairs
$M_r(L)$	Ranking monotonicity of the ranking sequence L
N_r	Overall count of nodes having identical rank r
$DM(L)$	Distinct metric of the ranking sequence L

The global influence is expressed for node p as:

$$GI(p) = KS(p) \tag{13}$$

The local influence is expressed for node p as:

$$LI(p) = I(p) + A(p) \tag{14}$$

where $I(p)$ indicates the personal influence of node p and $A(p)$ indicates the influence of its closest nodes on it.

The $I(p)$ is expressed as:

$$I(p) = D(p) \tag{15}$$

The $A(p)$ is expressed as:

$$A(p) = \frac{SUM(p)}{\max D} \tag{16}$$

where $SUM(p)$ stands for the total aggregate of the influence of the closest nodes of node p, and $\max D$ denotes the maximal node degree.

The $SUM(p)$ is expressed as:

$$SUM(p) = \sum\nolimits_{q\epsilon\Gamma(p)} \Pi(q) \tag{17}$$

where $\Pi(q)$ indicates the influence of closest node q, $\Gamma(p)$ indicates the set of closest nodes of node p, and $\Pi(q)$ can be expressed as:

$$\Pi(q) = D(q) \times Jacc(p, q) + KS(q) \tag{18}$$

where $Jacc(p, q)$ indicates the Jaccard similarity coefficient of node p and node q and is represented as:

$$Jacc(p, q) = \frac{|n(p) \cap n(q)|}{|n(p) \cup n(q)|} \tag{19}$$

where $n(p)$ and $n(q)$ indicate the set of nodes that contain links with node p and node q, respectively.

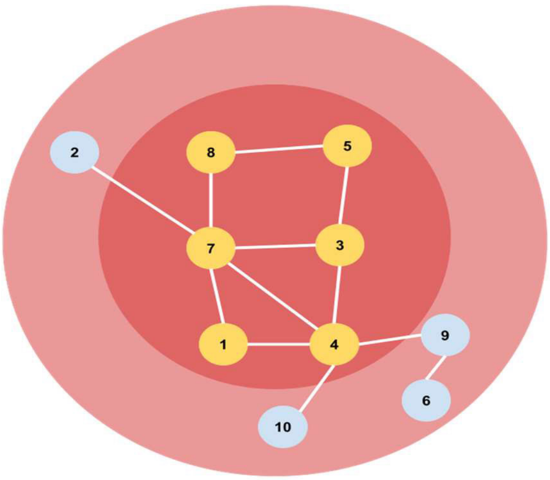

FIGURE 1. **A toy network with 10 nodes and 12 edges.**

IV. THE PROPOSED MODEL

We propose the degree k-shell eigenvector (DKE) index, which consists of three different components: degree, improved k-shell measure, and eigenvector centrality. Also, we propose an enhanced gravity model called the DKE-based gravity model (DKEGM) for discovering crucial nodes.

For the better clarity of the concept of the DKEGM, we use a simple toy network that contains 10 nodes and 12 edges, as displayed in Fig. 1. The values of degree, k-shell, improved k-shell, and eigenvector measures for each node within a toy network are indicated in Table 3. According to the Table 3, we can observe that $D(2) = D(6) = D(10) = 1$, $D(1) = D(5) = D(8) = D(9) = 2$, $D(4) = D(7) = 5$, $KS(2) = KS(6) = KS(9) = KS(10) = 1$, $KS(1) = KS(3) = KS(4) = KS(5) = KS(7) = KS(8) = 2$. So, the degree and k-shell measures both assign the identical value to the vast count of nodes, which creates a problem in distinguishing

TABLE 3. The values of degree, k-shell, improved k-shell, and eigenvector measures for each node within a toy network.

Node	1	2	3	4	5	6	7	8	9	10
D	2	1	3	5	2	1	5	2	2	1
KS	2	1	2	2	2	1	2	2	1	1
KS*	2.3333	1.3333	2.6667	2.6667	2.3333	1.3333	2.6667	2.3333	1.6667	1.3333
EC	0.3333	0.1688	0.3980	0.5142	0.2021	0.0587	0.5274	0.2335	0.1833	0.1646

TABLE 4. The values of DK and DKE indexes for each node within a toy network.

Node	1	2	3	4	5	6	7	8	9	10
DK	4.3333	2.3333	5.6667	7.6667	4.3333	2.3333	7.6667	4.3333	3.6667	2.3333
DKE	4.6666	2.5021	6.0647	8.1809	4.5354	2.3920	8.1941	4.5668	3.8500	2.4979

TABLE 5. The outcome of DKEGM ($R = 2$) for a toy network.

Node	1	2	3	4	5	6	7	8	9	10
1-order neighbors	4, 7	7	4, 5, 7	1, 3, 7, 9, 10	3, 8	9	1, 2, 3, 4, 8	5, 7	4, 6	4
2-order neighbors	2, 3, 8, 9, 10	1, 3, 4, 8	1, 2, 8, 9, 10	2, 5, 6, 8	4, 7	4	5, 9, 10	1, 2, 3, 4	1, 3, 7, 10	1, 3, 7, 9
DKEGM	99.14	35.19	154.23	235.38	66.79	14.10	235.18	82.58	61.33	34.66

these nodes. If we consider the combination of both degree and k-shell measures, i.e., (D + KS), the above-mentioned problem is still not solved. From Table 3, take the example that nodes 2, 6, and 10 have the same value of (D + KS), that is, 2. Also, nodes 1, 5, and 8 have the same value, that is, 4, and after that, nodes 4 and 7 have the same value, that is, 7. Also, if we consider the combination of both degree and improved k-shell measures, i.e., the DK index, we still have a problem in distinguishing these nodes. From Table 4, we can observe that $DK(2) = DK(6) = DK(10) = 2.3333$, $DK(1) = DK(5) = DK(8) = 4.3333$, $DK(4) = DK(7) = 7.6667$.

Therefore, to overcome the poor distinguishing ability of the degree, k-shell, combination of degree and k-shell measures (D + KS), and the DK index, we propose the degree k-shell eigenvector (DKE) index. The *DKE* index for node *p* is expressed as:

$$DKE(p) = D(p) + KS^{*}(p) + EC(p) \tag{20}$$

The value of the DKE index for every node within a toy network is displayed in Table 4, and we can see that each and every node within a toy network has a distinct value of the DKE index. Therefore, the DKE index has a high distinguishing ability.

According to the gravity law, we consider a node's DKE index value to represent its mass and the shortest path between two nodes to represent their distance. Thus, the influence of node *p* is expressed as:

$$DKEGM(p) = \sum\nolimits_{d(p,q)\leq R, q\neq p} \frac{DKE(p)DKE(q)}{d^{2}(p,q)} \tag{21}$$

where $DKE(p)$ and $DKE(q)$, respectively, stand for the DKE index values of nodes *p* and *q*. Such an enhanced gravity model is called the DKE-based gravity model (DKEGM). The enhanced gravity model depends on two facts: first, a node's influence increases if its neighbors have high values of the DKE index; second, the influence of a node on its neighbors decreases with the increase in shortest distances among them. The outcome of DKEGM ($R = 2$) for a toy network is given in Table 5. Consider an example of node 8, the neighbors of 1-order of node 8 are nodes 5 and 7, the neighbors of 2-order of node 8 are nodes 1, 2, 3, and 4. Therefore, $DKEGM\ (8) = DKE\ (8) * DKE\ (5) + DKE\ (8) * DKE\ (7) + DKE\ (8) * DKE\ (1)/4 + DKE\ (8) * DKE\ (2)/4 + DKE\ (8) * DKE\ (3)/4 + DKE\ (8) * DKE\ (4)/4 \approx 82.58$. Algorithm 1 provides an explanation of the DKEGM algorithm.

V. EXPERIMENTATION AND RESULTS' ANALYSIS

In this section, we show the outcomes of three different experiments executed by utilizing eight real networks of various sizes to evaluate the effectiveness of the DKEGM with respect to the other measures. Before displaying the empirical outcomes, we first introduce the experimental setup, real networks, and evaluation metrics used in the experiments.

TABLE 6. The structural attributes of eight real networks. These attributes include count of edges M, count of nodes N, average degree $< K >$, maximum degree K_{max}, infection threshold β_{th}, average clustering coefficient $< CC >$, average shortest distance $< d >$, and density σ.

Network	M	N	$< K >$	K_{max}	β_{th}	$< CC >$	$< d >$	σ
Dolphins	159	62	5.1290	12	0.1723	0.2590	3.3570	0.0841
Jazz	2742	198	27.6970	100	0.0266	0.6175	2.2350	0.1406
USAir	2126	332	12.8072	139	0.0231	0.6252	2.7381	0.0387
Netscience	914	379	4.8232	34	0.1424	0.7412	6.0419	0.0128
Wiki	2914	889	6.5557	102	0.0584	0.1528	4.0962	0.0074
Router	6632	2113	6.2773	109	0.0485	0.2464	4.6074	0.0030
Web-spam	37375	4767	15.6807	477	0.0140	0.2860	3.7935	0.0033
LastFM	27806	7624	7.2943	216	0.0409	0.2194	5.2322	0.0010

Algorithm 1 DKEGM

Input: Graph: $G = (V, E)$, Truncation radius: R, Count of nodes: N

```
Begin
for p ← 1toN do
    Compute D(p) using (7)
    Compute KS*(p) using (11)
    Compute EC(p) using (4)
    Compute DKE(p) using (20)
end for
for p ← 1toN do
    Find all neighbors of node p within the truncation radius R
    Compute DKEGM(p) using (21)
end for
Rank the influence of all nodes
return DKEGM(p)
End
```

Output: Sorted $DKEGM(p)$

A. EXPERIMENTAL SETUP

To evaluate the effectiveness of the DKEGM, we have executed the DKEGM and other nine measures in Python 3.7 using the NetworkX library and performed three different experiments on a system with the following configuration: CPU: AMD Ryzen 5 5600U, 6 cores, and 12 threads; Operating System: Windows 11; Memory: 16 GB; SSD: 512 GB. We have considered three parameters in the experimental setup, as follows:

1) ACCURACY

This reflects how the two ranking sequences, the real influence ranking sequence and the measure ranking sequence, are correlated [26].

2) DISTINGUISHING ABILITY

This indicates the ability of the centrality measure to distinguish the nodes' influence [26].

3) EFFICIENCY

This indicates the runtime of the centrality measure to compute the nodes' influence [26].

B. DATA DESCRIPTION

To examine the effectiveness of the DKEGM, we use eight real networks with different structural attributes. These networks come from several fields. The structural attributes of the eight networks are mentioned in Table 6. We now present these networks in brief, which are referred to at "https://networkrepository.com/networks.php" (Dolphins, Jazz, USAir, Netscience, Wiki, Router, Web-spam) and "https://snap.stanford.edu/data/feather-lastfm-social.html" (LastFM).

1) DOLPHINS [37]

This is a bottlenose dolphin social network. Every edge denotes a link between two dolphins, and every node denotes a bottlenose dolphin.

2) JAZZ [38]

Every node in this network of jazz musicians symbolizes a different musician, and every edge shows that two musicians are cooperating.

3) USAIR [39]

A network of air routes, where each airport serves as a node and each flight linking two airports as an edge.

4) NETSCIENCE [40]

This network consists of co-authorship links between researchers or scientists.

5) WIKI [41]

It is a network of Wikipedia voting information, where Wikipedia users are the nodes, and each edge linking nodes x and y indicates that user x cast a vote for user y.

6) ROUTER [41]

This is a technological network in which routers represent nodes, and each edge indicates a link connecting two routers.

7) WEB-SPAM [42]

This network is the Purdue University network repository, where nodes indicate web-pages and edges indicate hyperlinks.

8) LASTFM [43]

This is a social network with LastFM members. The nodes in the network represent members from Asia, while the edges show their mutual follower connections.

C. EVALUATION METRICS

SIR model [44], [45] is utilized to execute the process of spreading and compute the spreading influence of ranked nodes. Every node in this model is in one of the following three distinct states: (i) Susceptible state includes healthy individuals who are susceptible to infection from others. (ii) Infected state includes those individuals who have the illness and may be able to disseminate it to others. (iii) Recovered state includes infected individuals who have been recovered and are unable to spread the infection to others or be re-infected by others. Initially, only node x is in the infected state, while the remaining nodes are in the susceptible state. The nodes in the infected state attempt to spread the disease with infection probability β to their neighbors in the susceptible state at each time step. If the nodes in the susceptible state are infected, they enter the infected state. Then, the nodes of the infected state shift to the recovered state with a recovery probability μ. The recovery probability μ is specified to be 1.0 without losing generality, meaning that each infected individual can only infect his neighbors once before recovering at the following time step [46]. Until there are no infected nodes left within a network, this process continues. The process of spreading will be terminated once all of the infected nodes are extinct, and then the system will reach its stationary condition. All the recovered nodes at the completion of this process estimate the influence of the node x that was initially infected. In order to verify the nodes' spreading influence using the SIR model, the infection probability β must not be excessively small or large [12]. If β is extremely small, the infection cannot propagate throughout the network, and hence, the spreading influence of individual nodes cannot be determined. Conversely, if β is excessively large, the infection can quickly spread throughout nearly the whole network, which makes it impossible to identify the spreading influences of individual nodes. Therefore, for each network, we first determine the infection threshold β_{th}, which is expressed as:

$$\beta_{th} = \frac{< K >}{< K^2 > - < K >} \tag{22}$$

where $< K >$ denotes the average degree and $< K^2 >$ denotes the second-order average degree [47].

Kendall tau τ [48], [49] is utilized to find out the correlation between the ranking sequence attained using distinct centralities and the ranking sequence attained utilizing the SIR model. Consider two ranking sequences S and T that are correlated and have equal nodes N, $S = (s_1, s_2, \ldots, s_N)$ and $T = (t_1, t_2, \ldots, t_N)$. Let (s_a, t_a) and (s_b, t_b) be any two pairs: Both pairs are considered concordant if $(s_a > s_b)$ and $(t_a > t_b)$ or $(s_a < s_b)$ and $(t_a < t_b)$. Both pairs are considered discordant if $(s_a > s_b)$ and $(t_a < t_b)$ or $(s_a < s_b)$ and $(t_a > t_b)$. If $s_a = s_b$ or $t_a = t_b$, the pair is not considered as concordant or discordant. Kendall tau τ between S and T is expressed as:

$$\tau(S, T) = \frac{n_c - n_d}{0.5N(N-1)} \tag{23}$$

where n_d implies the overall count of discordant pairs, and n_c implies the overall count of concordant pairs. A higher value of τ shows a more accurate ranking sequence produced by a measure of centrality [50].

Ranking monotonicity M_r [20] is used to determine how distinctively different each element in a ranking list is, and it is expressed as:

$$M_r(L) = \left[1 - \frac{\sum_{r \epsilon L} N_r(N_r - 1)}{N(N-1)}\right]^2 \tag{24}$$

where L stands for the ranking sequence, N_r stands for the overall count of nodes having identical rank r. The range of M_r lies between 0 and 1. The measure has superior distinguishing ability when the M_r value is higher.

Distinct metric *DM* [51] is also utilized to measure the distinguishing ability of the centrality measures. It is expressed as:

$$DM(L) = \frac{number\ of\ nodes\ with\ distinct\ ranks}{N} \tag{25}$$

where L indicates the ranking sequence. The range of DM values lie between 0 and 1. The measure has better distinguishing ability if the value of DM is high.

D. PERFORMANCE EVALUATION

We have used four evaluation metrics, as mentioned above, to determine the performance of the DKEGM in finding the crucial nodes. First, we take a simple toy network that has 10 nodes and 12 edges, as shown in Fig. 1. We compare the DKEGM with other centralities, including DC, BC, CC, EC, KS, GLI, GIC, LGM, and DKGM, in terms of Kendall tau τ for the toy network, as displayed in Fig. 2. Fig. 2 shows that DKEGM performs best as compared to other centrality measures in terms of Kendall tau τ. The overall experimentation is categorized into the following three experiments on eight real networks, which are explained below:

1) COMPARE THE ACCURACY OF DIFFERENT MEASURES OF CENTRALITY

In this experimentation, we investigate the accuracy of distinct centralities. For this purpose, we utilize the Kendall tau

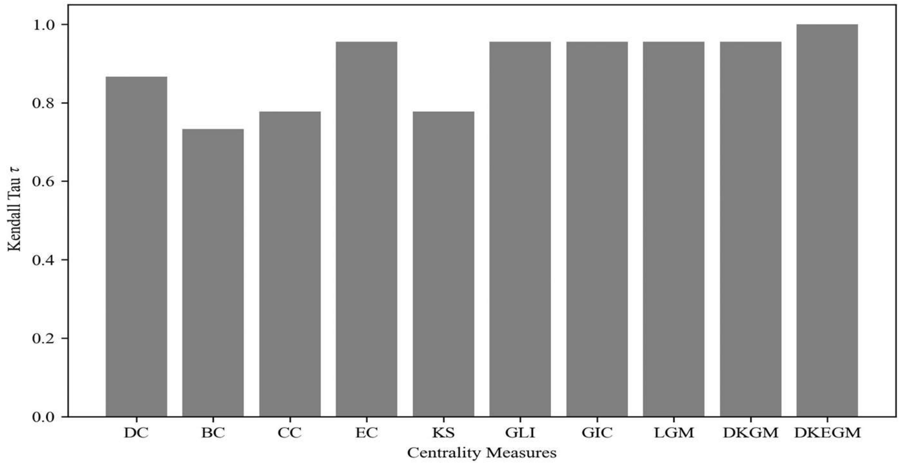

FIGURE 2. Kendall tau τ results of the toy network.

TABLE 7. The accuracy of the DKEGM and other centrality measures calculated by Kendall tau τ with $\beta = \beta_{th}$ on eight real networks.

Network	DC	BC	CC	EC	KS	GLI	GIC	LGM	DKGM	DKEGM
Dolphins	0.7472	0.5410	0.6584	0.6986	0.5219	0.7620	0.8488	0.8678	0.8900	**0.8964**
Jazz	0.8211	0.4688	0.7186	0.8913	0.7826	0.8421	0.8691	0.8688	0.8800	**0.8938**
USAir	0.7190	0.4975	0.7992	0.8809	0.7307	0.7934	0.8511	0.8798	0.8806	**0.8907**
Netscience	0.6099	0.3925	0.3403	0.4214	0.5704	0.6420	0.8082	0.8332	0.8443	**0.8461**
Wiki	0.6564	0.5017	0.7877	0.8263	0.6848	0.7545	0.8727	0.8643	0.8759	**0.8763**
Router	0.6356	0.4731	0.8102	0.8342	0.6610	0.7454	0.8485	0.8490	0.8585	**0.8587**
Web-spam	0.7124	0.5132	0.8093	0.8456	0.7315	0.7611	0.8252	0.8527	0.8578	**0.8578**
LastFM	0.5829	0.4112	0.6821	0.6761	0.6101	0.6595	0.8081	0.7771	0.7913	**0.8128**

τ to contrast the accuracy of the DKEGM with other measures of centrality. The DKEGM's truncation radius R is adjusted to its optimal value R^* according to the highest value of the Kendall tau [15]. Table 7 represents the τ values of the DKEGM and the other centrality measures with $\beta = \beta_{th}$ on eight real networks. The outcomes stated in Table 7 indicate the superiority of the gravity models (GIC, LGM, DKGM, and DKEGM) over the neighborhood-based (DC, KS, and GLI) as well as path-based (BC and CC) measures in all networks. As we can see from Table 6, the density of each network shows the sparse nature of the networks. The density of the network is the ratio of the count of actual connections to the total count of possible connections inside the network, which ranges from 0 to 1. A sparse network has a value closer to 0, and a dense network has a value closer to 1 [52].

Neighborhood-based (DC, KS, and GLI) as well as path-based (BC and CC) centralities do not perform well in sparse networks because neighborhood-based centralities encounter difficulties because of the few direct connections between nodes, while path-based centralities suffer from inefficiency when multiple nodes are needed to establish the connections between nodes in such networks, where direct connections may not exit. Gravity models (GIC, LGM, DKGM, and DKEGM) contain both neighborhood as well as path information, which enables better performance in sparse networks [14]. Also, EC performs well in some of the sparse networks because it incorporates the number of connections of a node as well as the importance of its connections. Furthermore, out of all gravity models, DKEGM provides the best performance in all networks because this model

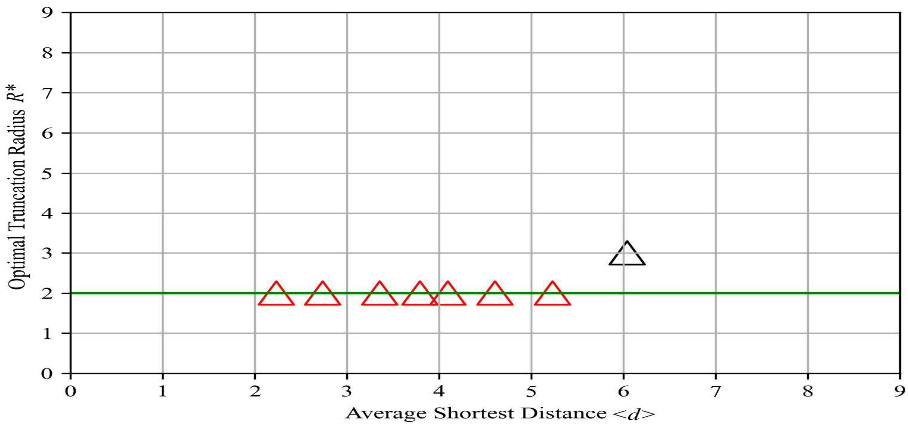

FIGURE 3. The relationship between optimal truncation radius R* of DKEGM and average shortest distance < d > for $\beta = \beta$th. Eight triangles represent eight networks and the green line represents $R = 2$. The black triangle is the Netscience network.

TABLE 8. The accuracy of DKEGM ($R = 2$) and other centrality measures calculated by Kendall tau τ with $\beta = \beta_{th}$ on eight real networks.

Network	DC	BC	CC	EC	KS	GLI	GIC	LGM	DKGM	DKEGM ($R = 2$)
Dolphins	0.7472	0.5410	0.6584	0.6986	0.5219	0.7620	0.8488	0.8678	0.8900	**0.8964**
Jazz	0.8211	0.4688	0.7186	0.8913	0.7826	0.8421	0.8691	0.8688	0.8800	**0.8938**
USAir	0.7190	0.4975	0.7992	0.8809	0.7307	0.7934	0.8511	0.8798	0.8806	**0.8907**
Netscience	0.6099	0.3925	0.3403	0.4214	0.5704	0.6420	0.8082	0.8332	**0.8443**	0.8356
Wiki	0.6564	0.5017	0.7877	0.8263	0.6848	0.7545	0.8727	0.8643	0.8759	**0.8763**
Router	0.6356	0.4731	0.8102	0.8342	0.6610	0.7454	0.8485	0.8490	0.8585	**0.8587**
Web-spam	0.7124	0.5132	0.8093	0.8456	0.7315	0.7611	0.8252	0.8527	0.8578	**0.8578**
LastFM	0.5829	0.4112	0.6821	0.6761	0.6101	0.6595	0.8081	0.7771	0.7913	**0.8128**

incorporates different aspects of nodes, which include count of neighbors, location of nodes, influence of neighbors, and path information between the nodes. Moreover, the BC measure exhibits poor performance, with the lowest values of correlation in most of the networks. The top measure that outperforms other measures is highlighted in bold, as given in Table 7.

Fig. 3 indicates the optimal values of the truncation radius R^* of the DKEGM for eight networks. We can observe that the optimal values of the truncation radius of the DKEGM for all networks, except the Netscience network, lie at $R = 2$. Also, most of the real networks exhibit small-world property [53], [54], and R^* needs to be small and can normally be fixed to 2 or 3. So, in order to save searching time for finding the optimal values of the truncation radius of the DKEGM for all networks, we can fix $R = 2$ to evaluate the effectiveness of the DKEGM. Table 8 compares the accuracy of the DKEGM for $R = 2$ with the other measures. As indicated in Table 8, the DKEGM for $R = 2$ shows excellent performance in all networks, excluding the Netscience network, where the DKGM attains the top position. In the Netscience network, the DKEGM does not attain the top position because its accuracy at $R = 2(\tau = 0.8356)$ is a little less as compared to its topmost accuracy at $R^* = 3(\tau = 0.8461)$. The top measure that outperforms other measures is highlighted in bold, as given in Table 8.

To further estimate how the variation in the value of β affects Kendall tau τ for different measures of centrality on

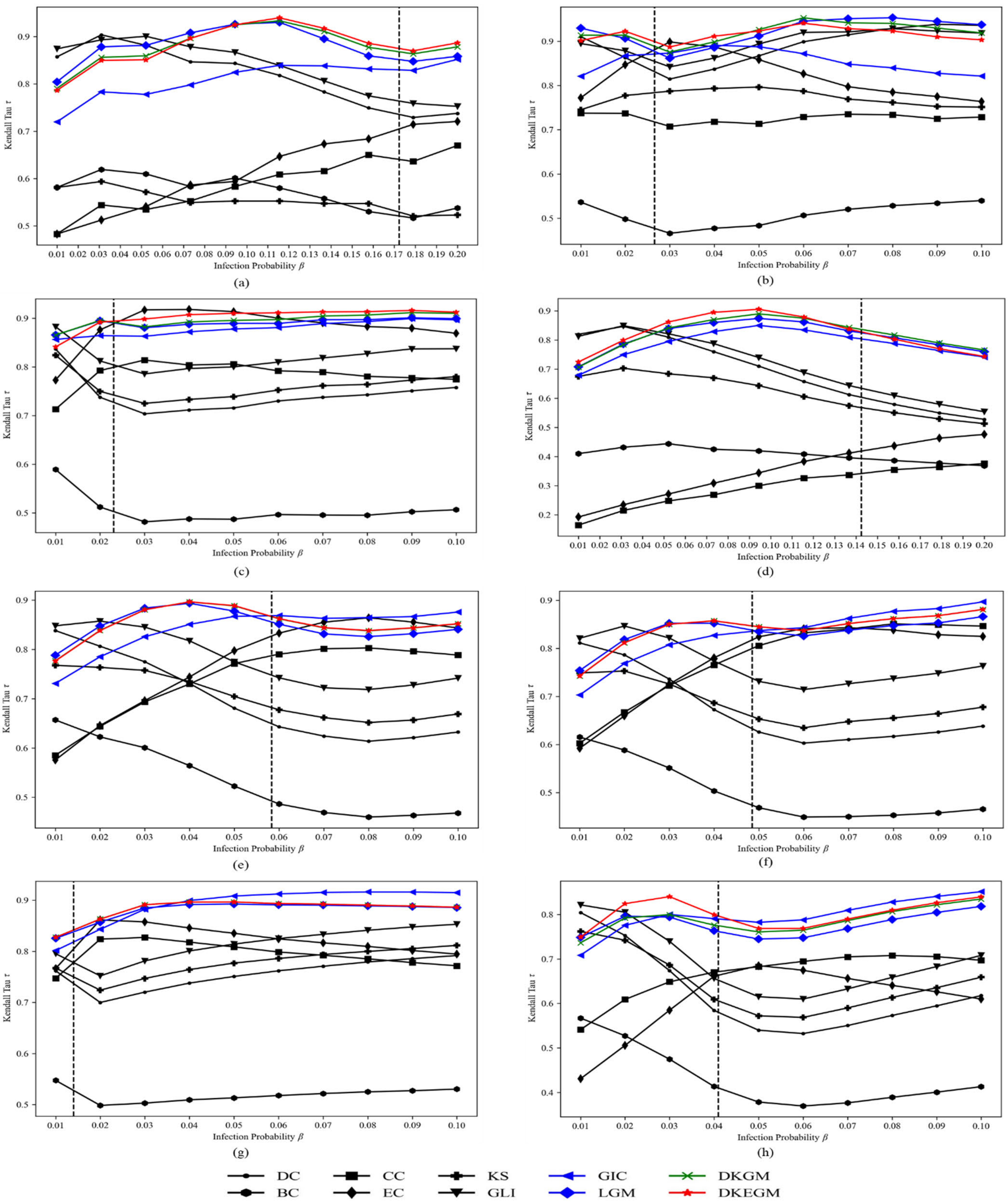

FIGURE 4. The accuracy of DKEGM and other benchmark centrality measures measured by Kendall tau τ for various infection probability β on eight real networks. (a) Dolphins. (b) Jazz. (c) USAir. (d) Netscience. (e) Wiki. (f) Router. (g) Web-spam. (h) LastFM. Dotted vertical line shows the infection threshold βth and the results were computed using the average result of 1000 independent runs.

TABLE 9. The value of ranking monotonicity M_r for ten measures of centrality on eight networks.

Network	DC	BC	CC	EC	KS	GLI	GIC	LGM	DKGM	DKEGM ($R = 2$)
Dolphins	0.8312	0.9623	0.9737	0.9979	0.3769	0.9905	0.9979	0.9937	0.9979	**0.9979**
Jazz	0.9659	0.9885	0.9878	0.9993	0.7944	0.9993	0.9993	0.9991	0.9993	**0.9993**
USAir	0.8586	0.6970	0.9892	0.9951	0.8114	0.9945	0.9943	0.9933	0.9940	**0.9956**
Netscience	0.7642	0.3390	0.9928	0.9952	0.6421	0.9941	0.9947	0.9950	**0.9957**	0.9955
Wiki	0.8041	0.8772	0.9988	0.9997	0.7265	0.9980	0.9997	0.9993	0.9998	**0.9998**
Router	0.7206	0.6691	0.9989	0.9992	0.6489	0.9963	0.9992	0.9984	0.9992	**0.9993**
Web-spam	0.8640	0.8471	0.9996	0.9998	0.8374	0.9992	0.9998	0.9998	**0.9999**	0.9998
LastFM	0.7978	0.8347	0.9997	0.9999	0.7348	0.9980	0.9998	0.9998	0.9998	**0.9999**

TABLE 10. The value of distinct metric *DM* for ten measures of centrality on eight networks.

Network	DC	BC	CC	EC	KS	GLI	GIC	LGM	DKGM	DKEGM ($R = 2$)
Dolphins	0.1935	0.8710	0.6935	0.9677	0.0645	0.9194	0.9677	0.9194	0.9677	**0.9677**
Jazz	0.3131	0.8939	0.6414	0.9646	0.1061	0.9646	0.9646	0.9545	0.9646	**0.9646**
USAir	0.1747	0.5542	0.5813	0.8313	0.0693	0.8223	0.8163	0.7861	0.8133	**0.8434**
Netscience	0.0554	0.2902	0.6016	0.7098	0.0211	0.6728	0.6887	0.6966	0.7230	**0.7230**
Wiki	0.0484	0.7334	0.7829	0.9561	0.0101	0.8571	0.9550	0.8763	0.9629	**0.9629**
Router	0.0293	0.5159	0.6924	0.8216	0.0071	0.7085	0.8202	0.6933	0.8230	**0.8410**
Web-spam	0.0365	0.6946	0.7258	0.9199	0.0073	0.8437	0.9159	0.8758	0.9226	**0.9241**
LastFM	0.0129	0.6763	0.7531	0.9227	0.0026	0.7636	0.9137	0.9180	0.9220	**0.9306**

eight real networks, see Fig. 4. According to Fig. 4, the DKEGM is the top performer among all the measures in all networks when β is not too far from β_{th}, except the Netscience network, where the DKGM shows the best outcome. Therefore, the ranking produced using the DKEGM becomes very close to the ranking produced using the SIR model, which we have also proved in Tables 7 and 8. The performance of the BC measure is very poor because of its smallest correlation values in most of the networks, as we have also proved in Table 7.

2) COMPARE THE DISTINGUISHING ABILITY OF DIFFERENT MEASURES OF CENTRALITY

In this experimentation, we investigate the distinguishing ability of distinct centralities. For this purpose, we utilize ranking monotonicity M_r to contrast the distinguishing ability of the DKEGM with the other measures of centrality. Table 9 represents the M_r values of the DKEGM and the other measures for eight networks. From Table 9, we have found that the DKEGM has higher M_r values in 6 out of 8 networks, i.e., Dolphins, Jazz, USAir, Wiki,

TABLE 11. The runtime (in seconds) of ten measures of centrality on eight networks.

Network	DC	BC	CC	EC	KS	GLI	GIC	LGM	DKGM	DKEGM ($R = 2$)
Dolphins	0.00004	0.0090	0.0035	0.0057	0.0002	0.0006	0.0147	0.0068	0.0108	0.0111
Jazz	0.0001	0.1843	0.0398	0.0220	0.0018	0.0159	0.3737	0.0545	0.0767	0.0884
USAir	0.0002	0.3456	0.0969	0.0107	0.0020	0.0141	0.8832	0.1178	0.1346	0.1012
Netscience	0.0003	0.3168	0.1207	0.0257	0.0011	0.0038	0.1274	0.1650	0.1769	0.0483
Wiki	0.0004	2.3416	0.7342	0.0407	0.0035	0.0142	2.6332	0.9289	0.9680	0.1973
Router	0.0011	14.3734	4.2361	0.1939	0.0084	0.0351	10.1158	5.0166	5.0985	0.5724
Web-spam	0.0026	140.6429	33.0498	0.2647	0.0473	0.3787	221.6140	43.3899	44.0772	7.0386
LastFM	0.0045	285.0151	67.5679	0.5685	0.0378	0.1652	74.0450	84.7651	85.8384	2.7330

Router, and LastFM, which indicates that the DKEGM is the best performer in terms of distinguishing ability, except in Netscience and Web-spam networks, where the DKGM performs well in terms of distinguishing ability. Moreover, the KS measure shows the worst performance in most of the networks.

We also utilize another metric called the distinct metric *DM* for further comparing the distinguishing ability of the DKEGM with the other measures of centrality. Table 10 represents the *DM* values of the DKEGM and the other measures for eight networks. From Table 10, it is observed that the DKEGM has higher *DM* values in all networks, which shows that the DKEGM is the best performer in terms of distinguishing ability. Moreover, the KS and DC measures show the worst performance in all networks. The top measure that outperforms other measures is highlighted in bold, as given in Tables 9 and 10.

3) COMPARE THE EFFICIENCY OF DIFFERENT MEASURES OF CENTRALITY

In this experimentation, we examine the runtime of distinct centralities to quantify the efficiency of each centrality in computing the influence of nodes. Table 11 compares the runtime of the proposed DKEGM with the runtime of the other centralities in eight real networks. As shown in Table 11, if we contrast the DKEGM with the other gravitational models (i.e., GIC, LGM, and DKGM) in terms of runtime, we note that the DKEGM has a low runtime in most of the networks. Therefore, this model can be used for large social networks. Also, the runtime of the DKEGM is low as compared to the runtime of the BC and CC measures in most of the networks. The runtime of the DC, EC, KS, and GLI measures is low as compared to the DKEGM, but at the same time, they do not have high accuracy as compared to the DKEGM in all the networks, as we have proved in Experiment 1. Therefore, overall, we can observe that the DKEGM is more effective in contrast with the other measures.

E. TIME COMPLEXITY

The DKEGM has four main components: degree, improved k-shell measure, eigenvector centrality, and R-order neighbors. The time complexity of the degree, improved k-shell measure, eigenvector centrality, and R-order neighbors of each node is $O(N)$, $O(M)$, $O(N + M)$, and $O(N < K >^R)$, respectively. So, the time complexity of our proposed model DKEGM can be $O(N + M + (N + M) + N < K >^R)$. The most complex operation of the DKEGM is computing each node's R-order neighbors, which has a time complexity of $O(N < K >^R)$. Hence, the overall time complexity of the DKEGM for achieving global information would be $O(N < K >^R)$. As we have observed in Fig. 3, the optimal values of the truncation radius of the DKEGM for all networks, except the Netscience network, lie at $R = 2$. Also, most of the real networks exhibit small-world property, and R^* needs to be small and can normally be fixed to 2 or 3. Hence, we set $R = 2$, which means the time complexity of the DKEGM is normally not greater than $O(N < K >^2)$, where $<K> \ll N$, which shows the DKEGM can be used for large and complex social networks. While the DKEGM's time complexity is not very high, it takes the network's global information into account when determining the nodes' influence.

VI. CONCLUSION

In this study, we proposed an index called the degree k-shell eigenvector (DKE) index that captures distinct aspects of a node's importance, which results in the high distinguishing ability of the proposed index. In addition, we proposed an enhanced gravity model called the DKE-based gravity model (DKEGM) for determining crucial nodes in social networks. The proposed gravity model incorporates the count of neighbors, location of nodes, influence of neighbors, and path information between the nodes. To evaluate the efficacy of the DKEGM, we executed different experiments on various real networks. Empirical outcomes indicated that the DKEGM has better accuracy and distinguishing ability as compared to other centrality measures. Moreover, the DKEGM has a low runtime in contrast to other gravity models. Therefore, this model can be used for large and complex social networks. Also, this model is suitable for unweighted and undirected networks and can be applied to both weighted and directed networks. In this study, we have analyzed the basic behavior of the DKEGM by assigning the same weights to the different measures for determining the crucial nodes in social networks. So, in our future work, we will apply the entropy weight method to the proposed model for assigning dynamic weights to the different measures.

REFERENCES

[1] A. Zareie and A. Sheikhahmadi, "A hierarchical approach for influential node ranking in complex social networks," *Exp. Syst. Appl.*, vol. 93, pp. 200–211, Mar. 2018.

[2] X. Huang, D. Chen, D. Wang, and T. Ren, "Identifying influencers in social networks," *Entropy*, vol. 22, no. 4, p. 450, Apr. 2020, doi: 10.3390/e22040450.

[3] A. Ullah, B. Wang, J. Sheng, and N. Khan, "Escape velocity centrality: Escape influence-based key nodes identification in complex networks," *Int. J. Speech Technol.*, vol. 52, no. 14, pp. 16586–16604, Nov. 2022, doi: 10.1007/s10489-022-03262-4.

[4] T. Bian and Y. Deng, "Identifying influential nodes in complex networks: A node information dimension approach," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 28, no. 4, Apr. 2018, Art. no. 043109, doi: 10.1063/1.5030894.

[5] M. J. Alvarez, Y. Shen, F. M. Giorgi, A. Lachmann, B. B. Ding, B. H. Ye, and A. Califano, "Functional characterization of somatic mutations in cancer using network-based inference of protein activity," *Nature Genet.*, vol. 48, no. 8, pp. 838–847, Aug. 2016, doi: 10.1038/ng.3593.

[6] Z. Yan-Chao, L. Yun, Z. Hai-Feng, C. Hui, and X. Fei, "The research of information dissemination model on online social network," *Acta Phys. Sinica*, vol. 60, no. 5, 2011, Art. no. 050501, doi: 10.7498/aps.60.050501.

[7] K. Das, S. Samanta, and M. Pal, "Study on centrality measures in social networks: A survey," *Social Netw. Anal. Mining*, vol. 8, no. 1, pp. 1–11, Dec. 01, 2018, doi: 10.1007/s13278-018-0493-2.

[8] P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," *J. Math. Sociol.*, vol. 2, no. 1, pp. 113–120, 1972.

[9] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, Dec. 1966, doi: 10.1007/BF02289527.

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

[11] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Phys.*, vol. 6, no. 11, pp. 888–893, Nov. 2010.

[12] L.-L. Ma, C. Ma, H.-F. Zhang, and B.-H. Wang, "Identifying influential spreaders in complex networks based on gravity formula," *Phys. A, Stat. Mech. Appl.*, vol. 451, pp. 205–212, Jul. 2016.

[13] A. Saxena and S. Iyengar, "Centrality measures in complex networks: A survey," 2020, *arXiv:2011.07190*.

[14] Z. Li and X. Huang, "Identifying influential spreaders in complex networks by an improved gravity model," *Sci. Rep.*, vol. 11, no. 1, Nov. 2021, Art. no. 22194, doi: 10.1038/s41598-021-01218-1.

[15] Z. Li, T. Ren, X. Ma, S. Liu, Y. Zhang, and T. Zhou, "Identifying influential spreaders by gravity model," *Sci. Rep.*, vol. 9, no. 1, p. 8387, Jun. 2019, doi: 10.1038/s41598-019-44930-9.

[16] A. Zeng and C.-J. Zhang, "Ranking spreaders by decomposing complex networks," *Phys. Lett. A*, vol. 377, no. 14, pp. 1031–1035, Jun. 2013.

[17] A.-L. Barabási and E. Bonabeau, "Scale-free networks," *Sci. Amer.*, vol. 288, no. 5, pp. 60–69, May 2003.

[18] P. Bonacich, "Some unique properties of eigenvector centrality," *Social Netw.*, vol. 29, no. 4, pp. 555–564, Oct. 2007.

[19] A. Bavelas, "Communication patterns in task-oriented groups," *J. Acoust. Soc. Am.*, vol. 22, no. 6, pp. 725–730, 1950.

[20] J. Bae and S. Kim, "Identifying and ranking influential spreaders in complex networks by neighborhood coreness," *Phys. A, Statist. Mech. Appl.*, vol. 395, pp. 549–559, Feb. 2014.

[21] K. Berahmand, A. Bouyer, and N. Samadi, "A new centrality measure based on the negative and positive effects of clustering coefficient for identifying influential spreaders in complex networks," *Chaos, Solitons Fractals*, vol. 110, pp. 41–54, May 2018.

[22] K. Berahmand, N. Samadi, and S. M. Sheikholeslami, "Effect of rich-club on diffusion in complex networks," *Int. J. Modern Phys. B*, vol. 32, no. 12, May 2018, Art. no. 1850142, doi: 10.1142/s0217979218501424.

[23] L. Fei, Q. Zhang, and Y. Deng, "Identifying influential nodes in complex networks based on the inverse-square law," *Phys. A, Stat. Mech. Appl.*, vol. 512, pp. 1044–1059, Dec. 2018.

[24] Z. Yu, J. Shao, Q. Yang, and Z. Sun, "ProfitLeader: Identifying leaders in networks with profit capacity," *World Wide Web*, vol. 22, no. 2, pp. 533–553, Mar. 2019, doi: 10.1007/s11280-018-0537-6.

[25] J. Dai, B. Wang, J. Sheng, Z. Sun, F. R. Khawaja, A. Ullah, D. A. Dejene, and G. Duan, "Identifying influential nodes in complex networks based on local neighbor contribution," *IEEE Access*, vol. 7, pp. 131719–131731, 2019, doi: 10.1109/ACCESS.2019.2939804.

[26] L. Qiu, J. Zhang, and X. Tian, "Ranking influential nodes in complex networks based on local and global structures," *Int. J. Speech Technol.*, vol. 51, no. 7, pp. 4394–4407, Jul. 2021.

[27] A. Ullah, B. Wang, J. Sheng, J. Long, N. Khan, and Z. Sun, "Identification of nodes influence based on global structure model in complex networks," *Sci. Rep.*, vol. 11, no. 1, p. 6173, Mar. 2021, doi: 10.1038/s41598-021-84684-x.

[28] X. Yang and F. Xiao, "An improved gravity model to identify influential nodes in complex networks based on k-shell method," *Knowl.-Based Syst.*, vol. 227, Sep. 2021, Art. no. 107198, doi: 10.1016/j.knosys.2021.107198.

[29] A. Ullah, B. Wang, J. Sheng, J. Long, N. Khan, and Z. Sun, "Identifying vital nodes from local and global perspectives in complex networks," *Expert Syst. Appl.*, vol. 186, Dec. 2021, Art. no. 115778, doi: 10.1016/j.eswa.2021.115778.

[30] H. Hu, Z. Sun, F. Wang, L. Zhang, and G. Wang, "Exploring influential nodes using global and local information," *Sci. Rep.*, vol. 12, no. 1, pp. 1–16, Dec. 2022, doi: 10.1038/s41598-022-26984-4.

[31] J. Liu and J. Zheng, "Identifying important nodes in complex networks based on extended degree and E-shell hierarchy decomposition," *Sci. Rep.*, vol. 13, no. 1, pp. 1–10, Feb. 2023.

[32] A. Ibnoulouafi and M. El Haziti, "Density centrality: Identifying influential nodes based on area density formula," *Chaos, Solitons Fractals*, vol. 114, pp. 69–80, Sep. 2018, doi: 10.1016/j.chaos.2018.06.022.

[33] J. Wang, C. Li, and C. Xia, "Improved centrality indicators to characterize the nodal spreading capability in complex networks," *Appl. Math. Comput.*, vol. 334, pp. 388–400, Oct. 2018.

[34] J. Sheng, J. Dai, B. Wang, G. Duan, J. Long, J. Zhang, K. Guan, S. Hu, L. Chen, and W. Guan, "Identifying influential nodes in complex networks based on global and local structure," *Phys. A, Stat. Mech. Appl.*, vol. 541, Mar. 2020, Art. no. 123262, doi: 10.1016/j.physa.2019.123262.

[35] H. Li, Q. Shang, and Y. Deng, "A generalized gravity model for influential spreaders identification in complex networks," *Chaos, Solitons Fractals*, vol. 143, Feb. 2021, Art. no. 110456, doi: 10.1016/j.chaos.2020.110456.

[36] Q. Shang, Y. Deng, and K. H. Cheong, "Identifying influential nodes in complex networks: Effective distance gravity model," *Inf. Sci.*, vol. 577, pp. 162–179, Oct. 2021.

[37] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, "The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations," *Behav. Ecol. Sociobiol.*, vol. 54, no. 4, pp. 396–405, Sep. 2003, doi: 10.1007/s00265-003-0651-y.

[38] P. M. Gleiser and L. Danon, "Community structure in jazz," *Adv. Complex Syst.*, vol. 6, no. 4, pp. 565–573, Dec. 2003.

[39] *North American Transportation Atlas Data (NORTAD).* Accessed: 1997. [Online]. Available: http://vlado.fmf.uni-lj.si/pub/networks/pajek/

[40] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 74, no. 3, pp. 1–19, Sep. 2006, doi: 10.1103/physreve.74.036104.

[41] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 4292–4293. [Online]. Available: http://networkrepository.com

[42] C. Castillo, K. Chellapilla, and L. Denoyer, "Web spam challenge," in *Proc. 4th Int. Work. Advers. Info. Ret. Web*, 2008, pp. 1–38.

[43] B. Rozemberczki and R. Sarkar, "Characteristic functions on graphs: Birds of a feather, from statistical descriptors to parametric models," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 1325–1334.

[44] L. J. S. Allen, "Some discrete-time Si, SIR, and SIS epidemic models," *Math. Biosci.*, vol. 124, no. 1, pp. 83–105, Nov. 1994.

[45] Y. Gang, Z. Tao, W. Jie, F. Zhong-Qian, and W. Bing-Hong, "Epidemic spread in weighted scale-free networks," *Chin. Phys. Lett.*, vol. 22, no. 2, pp. 510–513, Feb. 2005.

[46] C. Li, L. Wang, S. Sun, and C. Xia, "Identification of influential spreaders based on classified neighbors in real-world complex networks," *Appl. Math. Comput.*, vol. 320, pp. 512–523, Mar. 2018, doi: 10.1016/j.amc.2017.10.001.

[47] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, "Epidemic outbreaks in complex heterogeneous networks," *Eur. Phys. J. B Condens. Matter Complex Syst.*, vol. 26, no. 4, pp. 521–529, Apr. 2002.

[48] M. G. Kendall, "The treatment of ties in ranking problems," *Biometrika*, vol. 33, no. 3, pp. 239–251, 1945.

[49] W. R. Knight, "A computer method for calculating Kendall's tau with ungrouped data," *J. Amer. Stat. Assoc.*, vol. 61, no. 314, pp. 436–439, Jun. 1966.

[50] A. Ullah, B. Wang, J. Sheng, J. Long, and N. Khan, "Identification of influential nodes via effective distance-based centrality mechanism in complex networks," *Complexity*, vol. 2021, pp. 1–16, Feb. 2021, doi: 10.1155/2021/8403738.

[51] S. Gao, J. Ma, Z. Chen, G. Wang, and C. Xing, "Ranking the spreading ability of nodes in complex networks based on local structure," *Phys. A, Stat. Mech. Appl.*, vol. 403, pp. 130–147, Jun. 2014.

[52] G. Bo, Z. Xiangzhen, X. Biyu, M. Runnian, Z. Fengyue, and L. Wenbin, "Detecting dense subgraphs in complex networks based on edge density coefficient," *Chin. J. Electron.*, vol. 22, no. 3, pp. 517–520, Jul. 2013.

[53] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.

[54] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, "Classes of small-world networks," *Proc. Nat. Acad. Sci. USA*, vol. 97, no. 21, pp. 11149–11152, Oct. 2000.

HARDEEP SINGH received the B.Tech. and M.Tech. degrees in computer science and engineering from Guru Nanak Dev University, Amritsar, Punjab, India, in 2010 and 2011, respectively, where he is currently pursuing the Ph.D. degree with the Computer Science Department.

Since 2012, he has been an Assistant Professor with the Computer Engineering and Technology Department, Guru Nanak Dev University. He has published more than 15 research articles. His research interests include software engineering, social network analysis, and data mining.

HARDEEP SINGH (Member, IEEE) received the M.S. degree in software systems from the Birla Institute of Technology and Science, Pilani, Rajasthan, India, in 1992, and the Ph.D. degree in computer science and engineering from Guru Nanak Dev University, Amritsar, Punjab, India, in 2003.

From 2003 to 2023, he was a Professor with the Computer Science Department, Guru Nanak Dev University. He is the author of four books and he has published more than 180 research articles. His research interests include software engineering and information systems.