scientific reports

Check for updates

OPEN

# Influential nodes identification for complex networks based on multi-feature fusion

Shaobao Li, Yiran Quan, Xiaoyuan Luo & Juan Wang

**dentifying critical nodes in complex networks presents a significant challenge that has garnered extensive research attention. Previous studies often overlook the importance of spatial information, thereby limiting the accurate identification of key nodes. To address this gap, we introduce an advanced centrality model, termed Degree-*k*-shell-Betweenness Centrality (DKBC), which is grounded in the principle of gravity. The DKBC model integrates the centrality attributes of node degree, spatial positioning, and intermediate degree, resulting in improved accuracy for key node identification in complex networks. This innovative approach outperforms traditional gravity-based methods in terms of effectiveness. We validated the diffusion capacity of the proposed model using the Susceptible-Infected-Recovered (SIR) epidemic model and the Independent Cascade (IC) model, assessing correlation through the Kendall coefficient $\tau$. A comparative analysis with benchmark algorithms highlights the superior performance of the DKBC model. Empirical validation across twelve real-world networks demonstrates the model's exceptional accuracy in identifying key nodes. This study significantly advances the field by illustrating the effectiveness of incorporating spatial information into centrality measures to enhance both network analysis and practical applications.**

**Keywords** Complex networks, Influential node identification, Betweenness centrality, *k*-shell algorithm, Gravity model

Complex networks have emerged as a powerful tool for investigating the structure and dynamics of diverse systems spanning biological, social, technological, and transportation domains. These networks consist of interconnected nodes that exhibit distinctive topological properties, which profoundly influence system behavior. By analyzing complex networks, researchers can gain insights into real-world complexity and uncertainty[1,2]. Certain key nodes play a pivotal role in complex systems and have a significant impact on various structural characteristics of the network[3–5]. Consequently, the identification of these critical nodes has become a focal point for researchers aiming to accurately uncover hidden features for complex networks. Identifying crucial nodes in network topology is fundamental in complex network analysis. These critical nodes typically exhibit centrality metrics that are essential for ensuring efficient information flow, network robustness, and overall performance. Several classical techniques have been developed to pinpoint these crucial nodes in complex networks, including betweenness centrality[6], closeness centrality[7], degree centrality[8], eigenvector centrality[9], and *k*-shell decomposition method[10]. Each of these methods captures a different aspect of node influence in the complex network: Betweenness centrality is influenced by a node position in shortest paths, closeness centrality by a node proximity to others, degree centrality by a node number of neighbors, eigenvector centrality by the roles of its neighboring nodes, and *k*-shell decomposition method by the central position of the node in the network.

However, traditional techniques for assessing node importance in complex networks have certain limitations. Degree centrality, for example, focuses primarily on local node connections and fails to consider the broader network structure[11]. Betweenness centrality and closeness centrality provide comprehensive node rankings but can be computationally intensive, especially in large complex networks[12]. Moreover, eigenvector centrality cannot be applied to weighted networks, limiting its applicability. In recent years, new methods have emerged to overcome these limitations. Kumar et al.[13] introduced NCVoteRank, which leverages a coreness-based voting mechanism to identify influential spreaders by considering the coreness values of neighboring nodes. While innovative, this method may face computational efficiency challenges, especially in large-scale networks where frequent updates to the network structure are required. Alshahrani et al.[14] proposed the "MinCDegKatz d-hops" algorithm, blending established centrality metrics for effective propagation while ensuring low operational complexity. However, it may not fully capture the multi-layered influence of nodes within the network. Wen

School of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China. email: juanwang@ysu.edu.cn

et al.[15] employed a multi-local dimension (MLD) approach, utilizing fractal characteristics to detect essential diffusers. This method may perform poorly when the network structure is regular or the fractal characteristics are not pronounced. Xiao et al.[16] developed a parallel algorithm for identifying influential nodes in extensive biological networks, capitalizing on the power of GPUs. While this approach enhances processing capabilities, it may require significant hardware resources and may not be suitable for environments with limited resources. Zareie et al.[17] presented algorithms that harness the diversity among a node's neighbors to determine its ranking value. This approach may struggle to accurately assess node importance when neighbor node information is insufficient or when neighbor nodes exhibit little variance in influence. Jalili et al.[18] introduced IMUD, an algorithm designed to pinpoint the most impactful group of users. However, it may have difficulty accurately locating key users in networks with complex structures or intersecting user groups. Sheikhahmadi et al.[19] used user interaction levels to gauge the strength of relationships within networks, aiming to identify key influencers in social media environments. This method may be affected by changes in user behavior patterns, resulting in unstable assessment of relationship strength.Hu et al.[20] focused on identifying influential communicators within community networks, emphasizing the crucial role of inter-community distances. However, this method may not fully account for the impact of network structures other than community-based organizations, limiting its application in networks that are not community-driven or have weak community structures. Bian et al.[21] introduced a node information dimension-based metric for identifying pivotal nodes, which may encounter difficulties when dealing with networks where node information dimensions are unclear or overlapping. Ren et al.[22] proposed a method to assess node propagation capacity using their minimum $k$-shell values, with particular emphasis on specific nodes, but this approach might overlook other significant nodes that are not central within the network. Zhang et al.[23] presented a technique for selecting transmission sources with optimal transmission capability, effectively mitigating the rich club phenomenon, but this method may struggle to adapt in networks with rapidly changing structures. Guo et al.[24] proposed an enhanced distance-based coloring technique aimed at identifying pivotal spreaders; despite its significant computational complexity, this method evaluates the local dimension across various topological distances for each node. Li et al.[25], inspired by the gravitational law, introduced the gravity centrality model, where node degrees act akin to mass, and the shortest path distances between nodes signify their separation. Node interactions are influenced by both their degrees and distances, mirroring physical interactions governed by gravity. The gravity centrality model implies a consistent attraction among nodes, where a node's influence is determined by its position and neighboring nodes[26]. However, this method may perform poorly in networks with uneven density or when node distance information is inaccurate, suggesting that nodes in lower $k$-shell layers can sometimes wield more influence than those in higher layers. While these methods have some validity in specific situations, their limitations become particularly apparent when dealing with dynamically changing or structurally complex networks. These limitations have led us to seek more accurate, robust and adaptable methods to identify key nodes in complex networks.

Combined with the previous discussion, this study focuses on the influence of node location and structure in complex networks, so as to adapt to networks with different community structures and achieve higher performance. It proposes that central nodes in a network have a greater ability to attract and influence other nodes than peripheral nodes[27]. Therefore, this paper introduces an enhanced gravity model called the degree-$K$-shell-intermediate centrality (DKBC) model, which incorporates multiple features. The model uses the concept of $k$-shell value to locate each node in the network as an attraction coefficient to adjust the influence of the central node. By integrating various node characteristics such as node path data, number of neighbors, mediation ability and node positioning, the influence of nodes in the propagation dynamics is evaluated. An empirical test is carried out on the real world network to prove the effectiveness of the proposed method. In brief, the contributions of this work are summarized as follows:

- A new method based on gravity model is proposed. The influence of a node is determined by both local and global information.
- Attraction coefficients are introduced to describe the attraction of nodes in different nuclear layers, and tunable parameters are set to adjust the algorithm to deal with networks with different structures.
- The proposed method shows good diffusion ability on different models and has superior accuracy and monotonicity.

The structure of this paper is as follows: "Related works" briefly reviews some relevant studies and presents relevant data and methods. "Methods" details the DKBC model. In "Data and experiments", we demonstrate the rationality of the introduced method through some real network practice. "Conclusions" is the summary of this paper.

## Related works

In the realm of complex network analysis, identifying influential spreaders is a pivotal task with broad applications across various domains. Numerous studies have been dedicated to developing centrality measures that can accurately pinpoint these key nodes. This section reviews the state-of-the-art in this field, highlighting the contributions and limitations of the existing approaches.

### Local based methods

Some centrality measures evaluate the propagation ability of nodes from local structure information, and hold that the influence of nodes is largely affected and reflected by the topology of the network they belong to. For example, the DCL method proposed by Berahmand et al.[28] which identifies key spreaders in social networks based on the degree of nodes, the degree of neighbors, the number of links between common neighbors, and the inverse local clustering coefficient. The principle of this method is to comprehensively consider the local network structure characteristics of nodes and evaluate the propagation ability of nodes from a multi-dimensional perspective. Its advantage is that it can efficiently identify important nodes in the network with nearly linear time complexity without adjusting any free parameters, showing high accuracy and applicability. However, the limitation of DCL method is that its local perspective may not be able to fully capture the overall structural features of the network, and for network regions with high clustering coefficients, the DCL value of nodes may be affected, which limits its applicability in some specific network structures. Bouyer et al.[29] proposed an ERND method based on Edge Ratio and Neighborhood Diversity. This method determines the influence of nodes as communicators by evaluating the edge ratio and neighborhood diversity of nodes. The edge ratio ensures that the selected nodes are not at the edge of the network. The neighborhood diversity indicates that the node may be the link between some modules in the network. The main advantages of the ERND approach are its near-linear time complexity, no parameter tuning, emphasis on global bridging nodes, and high accuracy using only local information. However, the limitation of this method is that it mainly relies on the local network structure and may not adequately capture the global characteristics of the network, which in some cases may affect the identification of global influence nodes.

### Global based methods

The primary goal of the aforementioned algorithms is to assess node significance based on local network information. However, in practice, each node's ability to influence traffic and information flow during dissemination significantly impacts its overall importance. From the perspective of spreading dynamics, a node's capacity to facilitate rapid and widespread information dissemination indicates its greater importance. Building on this insight, various methods have been proposed to rank nodal significance by utilizing global network information. Zhong et al.[30] proposed an innovative algorithm termed Local Degree Dimension, which evaluates the propagation capacity of nodes by examining the growth and decline rates of neighbor counts across different layers of central nodes. In a similar vein, Ullah et al.[31] developed a centrality measure that incorporates both local and global perspectives to effectively identify the most significant nodes within a network, addressing both local and global topological features simultaneously. Ruan et al.[32] proposed a gravity-based method for evaluating node importance in complex networks (ISM and its extended algorithm ISM+). This method comprehensively considers the H-index, K-core centrality and structural hole characteristics of nodes. It is a method that synthesizes local and global information, can integrate multiple attribute information, and improves the comprehensiveness and accuracy of node importance assessment. However, it mainly considers the shortest path between nodes and ignores other possible path information, which may affect the complete evaluation of the interaction effect between nodes. In addition, community perception centrality measurement is also an important method in the field of key node identification in complex networks. Savonnet et al.[33] emphasizes the effectiveness of community perception centrality measurement in considering the structure of network communities. Through comparative analysis, this study reveals the correlation between different measures of community perceived centrality and explores their effects on propagation in the SIR (susceptible-infected-recovered) model. On this basis, Rajeh et al.[34] further discusses the impact of community structure on the centrality measure, emphasizing that in strong community structure networks, the correlation between the community perceived centrality measure and the traditional centrality measure is low, while in weak community structure networks, the correlation is high. Community perceived centrality measurement has superior performance on networks with distinct community structure.

### Gravitational model-based methods

In recent years, the method of identifying important nodes based on the improved gravity model has also received great attention[35–39]. The application of gravity model in complex network analysis is inspired by the law of gravitation in physics. In this model, nodes are treated as objects with mass, and the edges between nodes are treated as the distances between them. The core idea of the model is that the importance of nodes depends not only on their own "quality" (such as degree centrality, *k*-shell value, etc.), but also on their position relationship in cyberspace. The advantage of the gravity model is that it can consider not only the local connections of nodes, but also the location of nodes in the global network, which allows it to assess the importance of nodes more comprehensively. In addition, strong adaptability is also an advantage of the gravity model. By adjusting the model parameters, the gravity model can adapt to different types of network structures and application scenarios, providing a multi-dimensional perspective for the identification of key nodes[35]. Liu et al.[36] introduced weighted gravity centrality, which combines eigenvector centrality with gravity model and integrates global and local information. However, weighted model requires reasonable setting of weight parameters, and parameter optimization may be a challenge. Yang[38] et al. introduced KSGC, which adds an attraction coefficient to the generalized gravity model based on *k*-shell. However, due to the limited information expressed by node degree, the use of node degree as mass generally lacks accuracy. Recently, Xu et al.[39] proposed a new adaptive gravity model based on communication(CAGM), which is an innovative adaptive gravity model based on communication. The model can dynamically adjust the interaction intensity between nodes according to the communication conditions of the network, improving the adaptability and accuracy of the model. However, the dynamic adjustment mechanism may introduce instability in the model, and further research is needed to balance the frequency of dynamic adjustment and model stability.

## Methods

### Centrality approach

A complex network can be expressed as $G = (V, E)$, in which $V$ denotes the collection of nodes and $E$ indicates the collection of edges. The relationships within this network are represented by an adjacency matrix $A = [a_{ij}]$. In this matrix, a value of $a_{ij} = 1$ signifies that there is a connection between nodes *i* and *j*, while a value of $a_{ij} = 0$ indicates the absence of any connection between them. Some important centrality measurement methods for complex networks are introduced in the sequel. The key symbols and abbreviations in this article are listed in Table 1.

Degree centrality (DC) is widely recognized as a fundamental measure of node significance in complex networks. Nodes possessing a higher degree are regarded as more significant due to their capacity to affect a greater number of neighboring nodes. The normalized DC can be expressed as:

$$DC(i) = \frac{k_i}{N-1} \tag{1}$$

In this context, $N$ refers to the total count of nodes in the graph $G$, while $k_i$ signifies the degree of node $i$, reflecting the quantity of neighbors to which it is directly connected.

Betweenness centrality (BC) serves as an additional metric for assessing the importance of nodes in a network. This measure calculates how often a node is present on the shortest paths connecting various pairs of other nodes, thereby emphasizing its function as a crucial linkage. The BC of node $i$ is defined as follows:

$$BC(i) = \frac{\sum_{x \neq y \neq i} L_{xy}(i)}{\sum_{x \neq y} L_{xy}} \tag{2}$$

where $L_{xy}$ denotes the number of shortest paths from node $x$ to node $y$, and $L_{xy}(i)$ denotes the number of shortest paths from node $x$ to node $y$ passing through node $i$.

Closeness centrality (CC) serves as a metric that is calculated by taking the reciprocal of the mean shortest path length from a specific node to every other node within the network. Nodes with high closeness centrality are able to quickly communicate and disseminate information, making them important connectors within the network. The closeness centrality for node $i$ can be determined using the following formula:

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

In this context, $d_{ij}$ denotes the minimal distance along the path from node $i$ to node $j$.

The concept of eigenvector centrality (EC) quantifies the strength of relationships between a node and its adjacent nodes. This indicates that the significance of a particular node is dependent on the properties of its neighbors. The effective connectivity for node $i$ is formally expressed as follows:

$$x(i) = c \sum_{j=1}^{N} a_{ij} x(j) \tag{4}$$

| Parameter | Description |
|---|---|
| $G$ | An unweighted and undirected network |
| $V$ | Collection of nodes in $G$ |
| $N$ | Number of nodes in $G$ |
| $E$ | Collection of edges in $G$ |
| $A$ | Adjacency matrix of $G$ |
| DC | Degree centrality |
| $k_i$ | Degree of node $i$ |
| $ks(i)$ | $K$-shell value of node $i$ |
| EC | Eigenvector centrality |
| BC | Betweenness centrality |
| CC | Closeness centrality |
| GC | Gravity centrality |
| $R^*$ | Ideal truncation radius |
| $d_{ij}$ | Minimal distance along the path from node $i$ to node $j$ |
| $\beta$ | Epidemic transmission probability in SIR Model |
| $\lambda$ | Recovery probability of an infected node in the SIR Model |
| $p_{uv}$ | Activation probability of nodes in IC model |
| $F(t)$ | The transmission capability of nodes in the SIR Model |
| $I(t)$ | The propagation ability of nodes in IC models |
| $c_{ij}$ | The attraction coefficient between node $i$ and node $j$ |
| $\varepsilon$ | Tunable parameters in the DKBC model |

**Table 1.** Notations and abbreviations appears in this paper.

In most cases, the constant $c$ is defined as the inverse of the largest eigenvalue of matrix $A$, which can be represented as $c$.

Gravity centrality (GC) is a measure that captures the strength and proximity of connections between nodes, analogous to the gravitational principle in physics. Nodes characterized by elevated gravity centrality serve as vital connections between various segments of the network, enhancing the transmission of information and impact. The evaluation of the gravity centrality for node $i$ can be expressed as follows:

$$GC(i) = \sum_{j \neq i, d_{ij} \leq R} \frac{k_i k_j}{d_{ij}^2} \tag{5}$$

In this paper, parameter $R$ represents the truncation radius, which is used to address the issue of lengthy large-scale network search. The subsequent formula can be utilized to determine the ideal truncation radius, denoted as $R^*$:

$$R^* \approx 0.5\langle d \rangle \tag{6}$$

In this context, $\langle d \rangle$ represents the mean distance within the network.

Inspired by the gravity model, Ruan et al.[32] proposed a new node importance evaluation method (ISM), which comprehensively considered the H-index of nodes, the number of cores and the location of structural holes to identify key nodes in the network. ISM is defined as:

$$ISM(i) = \sum_{d_{ij} \in \psi_i} e^{-C_i} \frac{(ks_i + \gamma h_i)(ks_j + \gamma h_j)}{2d_{ij}^2} \tag{7}$$

where $\psi_i$ is the set of neighborhood nodes whose distance from node $i$ is less than or equal to the given value r, $ks_i$ is the $k$-shell value of node $i$. $C_i$ represents the constraints under which the node forms a structural hole, $\gamma$ represents the equilibrium factor, and $h_i$ represents the H-index of node $i$. In order to reduce the computational complexity, the r value is set to 3, and further, the extension of the ISM method ISM+ is defined as:

$$ISM^+(i) = \sum_{j \in \Gamma_i} ISM(j)^\theta \tag{8}$$

where $\Gamma_i$ represents the set of neighbors of node $i$, $0 \leq \theta \leq 1$. For smaller $\theta$, the ISM+ method weakens the influence of influential neighbors with larger ISM values, while larger $\theta$ values enhance the influence of influential neighbors with larger ISM values. Without loss of generality, $\theta$ was taken as 0.8 in subsequent experiments.

Ibnoulouafi et al.[40] proposed M-Centrality (MC), which effectively combines the global position and local degree variation of nodes. MC is defined as:

$$M(i) = \mu ks_i + (1 - \mu)\Delta D_i \tag{9}$$

where $\mu$ is a weight factor determined by Shannon entropy method, which is used to balance the importance of global position information and local degree change information. $\Delta D_i$ is the degree change in the neighborhood of node $i$ and quantifies the degree distribution in the local neighborhood of the node. $\Delta D_i$ is defined as:

$$\Delta D_i = \sum\nolimits_{j \in \Gamma_i} |k_j - k_i| \tag{10}$$

where $k_j$ and $k_i$ are the degrees of nodes $j$ and $i$ respectively.

Based on the inter-network Community perception, Ghalmane et al.[41] proposed the Community Hub-Bridge (CHB) method, which identifies key nodes in the network by combining the connection of nodes within the community (as “Hub”) and the connection across the community (as “Bridge”). Suitable for networks with a distinct community structure that captures both the local and global influence of nodes. CHB is defined as:

$$CHB(i) = \frac{n_{C_q} \cdot k_i^{\text{intra}}}{N} + \frac{k_i^{\text{inter}}}{C} \tag{11}$$

where $n_{C_q}$ is the size of the community where node $i$ resides, $k_i^{\text{intra}}$ is the number of connections of node $i$ within the community, $N$ is the total number of nodes in the network, $k_i^{\text{inter}}$ is the number of connections of node $i$ across communities, and $C$ is the total number of communities in the network.

### K-shell algorithm

In complex network analysis, the $k$-shell technique serves as a method to categorize nodes in a network, uncovering their hierarchical roles and significance. This approach offers a deeper understanding of the connections and importance of individual nodes within the network structure. A $k$-shell is defined as the set of nodes belonging to the largest subgraph in which all nodes have a degree of at least $k$. At the outset of the algorithm, the degree of each node is assessed, and the process begins by eliminating all nodes that exhibit a degree of $k = 1$, as these are considered to have moderate connectivity. Subsequent rounds of removal target nodes with degrees of $k \leq 1$, systematically diminishing the moderate degree values during the process. This iterative procedure persists until there are no nodes with a degree of $k \leq 1$ left in the network. The nodes that are removed in this stage constitute the 1-shell, which is characterized by a *k*-shell value of 1. The algorithm is then repeated to create 2-shells, 3-shells, and so on, categorizing all nodes into distinct shells and assigning each node its corresponding *k*-shell value. By partitioning the network into different *k*-shells, we can identify the hierarchy in the network, as nodes with similar properties tend to gather together in different shells.

## The spreading model

*SIR model*

The key factor in assessing the significance of a node is its ability to disseminate information. Currently, there are several information propagation models available, such as the linear threshold model, disease transmission model, and independent cascade model. Among these, the SIR model is the most widely utilized. Numerous scholars utilize the SIR model to assess the magnitude of information and virus transmission. Consequently, this research employs the SIR framework to investigate the potential of network nodes for information dissemination. In the context of the SIR model, nodes are classified into one of three specific categories:

(I) Susceptible (S) nodes, which are at risk of infection yet remain uninfected;

(II) Infected (I) nodes, indicating nodes that have contracted the infection with a probability denoted as $\beta$;

(III) Recovered (R) nodes, representing nodes that were infected previously but have since recuperated, rendering them immune to future infections.

In the SIR model, all nodes are initially classified as being in the Susceptible (S) state, with the exception of the source node, which occupies the Infected (I) state. During each time interval, nodes that are infected have the potential to spread the infection to their neighboring susceptible nodes, with the likelihood of transmission indicated by the parameter $\beta$. Subsequently, these infected nodes can move to the Recovered (R) state with a probability denoted as $\lambda$. We use $F(t)$ to represent the number of recovered nodes in the network at a particular time t. This cycle of infection and recovery persists until the system attains a stable equilibrium. For ease of calculation, if the recovery rate $\lambda$ is assigned a value of 1, the critical threshold for the infectious disease can be determined using the subsequent formula:

$$\beta_c \approx \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \tag{12}$$

In this context, $\langle k \rangle$ signifies the mean degree, while $\left\langle k^2 \right\rangle$ indicates the second-order moment associated with the degree distribution.

*IC model*

The Independent Cascade Model (IC model) is a network diffusion model popular in sociology, communication, and computer science[42]. The model is widely used to simulate the spread of information, influence, or any other "diffuse" entity in a social network. In the IC model, each node can be in one of two states: Activated or Inactive. The model starts with one or more initial active nodes. At each time step, the active node attempts to activate its inactive neighbor node. For each pair of nodes ($u$, $v$), where u is an active node and v is an inactive node, there is a preset activation probability $p_{uv}$, which represents the probability that u can successfully activate v. The key assumption of the IC model is "independence", that is, the activation attempt of one node to another node is independent and not influenced by the other nodes. This process is iterated until no new nodes can be activated.

## Evaluation metrics

*Spreading ability*

As mentioned above, nodes with greater importance will have greater ability to propagate in the network. Therefore, in the SIR Model, in order to compare the propagation ability of high-ranked nodes obtained by different methods, we take the high-ranked nodes obtained by different methods as the initial infected nodes, and use $F(t)$ to represent the number of recovered nodes in the t network at a specific time. $F(t)$ increases with the infection process t and eventually reaches a stable value when there are no new recovery nodes in the network. In each method, we select the top five percent of nodes as the initial infected nodes. And the $F(t)$ could be obtained as:

$$F(t) = N_i(R) \tag{13}$$

where $N_i(R)$ is the number of recovery nodes at time t caused by the initial infected node whose ranking index is $i$. Each experiment was performed 1000 times in the corresponding network, where $\beta = \beta_c$. The final $F(t)$ represents the average result of the transmission capacity.

Similarly, in the IC model, in order to compare the propagation ability of higher-order nodes obtained by different methods, we take the higher-order nodes obtained by different methods as the initial active nodes, and the activation probability$p_{uv}$ is set separately for each node according to different centrality measures. Nodes with higher centrality measures have higher activation probabilities and are more likely to activate their neighbors. Activation probability $p_{uv}$ can be expressed as:

$$p_{uv} = \frac{CM(i)}{\max\{CM(i)\}} \tag{14}$$

where *CM*(*i*) represents the centrality measure of the node, and $\max\{CM(i)\}$ represents the largest centrality measure in the network. *I*(*t*) represents the number of active nodes in the network at a given time, and *I*(*t*) increases as the activation process t increases, eventually reaching a stable value when there are no new active nodes in the network. In each method, we select the top ten percent of the nodes as the initial active nodes. And *I*(*t*) is defined as:

$$I(t) = N_i(A) \tag{15}$$

where $N_i(A)$ is the number of active nodes at time t caused by the initial active node whose ranking index is *i*. Each experiment is conducted 1000 times in the corresponding network, and the final *I*(*t*) indicates the average result of influence.

*Kendall's Tau*
Kendall's Tau is a robust statistical measure that quantifies the degree of association between two ranked variables, revealing the extent to which their orderings are concordant or discordant. This metric is particularly insightful for analyzing ranked data in various fields, including social network analysis, where the ranking of nodes by different centrality measures can provide valuable insights into network structure and dynamics. The Kendall's Tau coefficient, ranging from -1 to 1, offers a nuanced view of the relationship between two datasets. A coefficient of 1 signifies a perfect positive correlation, indicating that the rankings of the two variables are identical. Conversely, a coefficient of -1 indicates a perfect negative correlation, where one variable's ranking is the exact reverse of the other. A value of 0 suggests that there is no monotonic relationship between the rankings, implying that the order of one variable provides no predictive power over the other. In the context of complex networks, Kendall's Tau serves as a critical tool for evaluating the consistency and reliability of different centrality measures. Let $X = (x_1, x_2, \ldots, x_m)$ and $Y = (y_1, y_2, \ldots, y_m)$ be two sequences, each comprising *m* elements, representing the rankings of nodes by two different measures within a network. The pairs of elements $(x_i, y_i)$ and $(x_j, y_j)$ are deemed concordant if they both exhibit the same relative ordering, either both increasing or both decreasing. Discordant pairs, on the other hand, exhibit opposite orderings. The absence of a tie in either variable results in a pair being neither concordant nor discordant. The calculation of Kendall's coefficient $\tau$ is given by:

$$\tau = \frac{2(\mathrm{m}_+ - m_-)}{m(m-1)} \tag{16}$$

here, $m_+$ represents the count of concordant pairs, and $m_-$ denotes the count of discordant pairs. This formula captures the net difference between agreements and disagreements in the rankings, normalized by the total number of possible pairs, providing a standardized measure of ranking correlation. The practical significance of Kendall's Tau in real-world networks cannot be overstated. It allows researchers and practitioners to assess the effectiveness of various centrality measures in identifying influential nodes, which is crucial for applications such as network marketing, disease control, and information dissemination. By quantifying the agreement between different centrality rankings, Kendall's Tau offers a means to validate, compare, and refine network analysis methodologies. This, in turn, enhances our ability to predict network behavior, design targeted interventions, and understand the underlying mechanisms that govern network dynamics. Furthermore, Kendall's Tau can reveal underlying structural properties of networks, such as the presence of hierarchical structures or the impact of community structures on node influence, providing deeper insights into the network's organization and function.

*Monotonicity*
In order to evaluate the precision of the ranking algorithm more comprehensively, we utilize the monotonicity index, represented as $M$, to gauge the monotonicity of the DKBC method and to contrast it with that of other comparative methods. The concept of ranking monotonicity is defined as follows:

$$M(x) = [1 - \frac{\sum_{r \in R} N_r(N_r - 1)}{N(N-1)}]^2 \tag{17}$$

In this context, $x$ signifies the sequence of nodes, $r$ refers to the ranking index, and $N_r$ represents the total number of nodes that possess the rank $r$. A value of $M(x)$ nearing 1 indicates a stronger level of monotonicity in the sequence, where each node is assigned a unique rank. Conversely, when $M(x)$ approaches 0, it suggests that the network contains only a single ranking value, meaning all nodes share the same rank.

## Proposed DKBC model

To sum up, we aim to combine the advantages of gravity model, propose a key node identification method with higher performance and can adapt to a variety of complex structure networks, and comprehensively consider the local and global information of nodes. The basic formula of the law of gravitation is:

$$F = G\frac{m_1 m_2}{r^2} \tag{18}$$

where *F* is the magnitude of the gravitational force between two objects, *G* is the universal gravitational constant, $m_1$ and $m_2$ are the masses of the two objects respectively, and *r* is the square of the distance between the two objects. Gravity models assess the interaction between node pairs to evaluate their influence. In this model, a node's influence is influenced by two primary factors. The first factor is the node's degree; a higher number of adjacent nodes suggests an increased capacity for direct influence, thus enhancing the overall effect of the node. The second factor involves the distance between nodes, as the strength of influence decreases in proportion to the square of the distance-nodes that are closer typically show more robust interactions. We assert that the influence one node exerts on another is intricately linked to their relative locations within the network, with nodes situated centrally demonstrating greater influence compared to those positioned at the edges. The importance of a node's location is captured by its $k$-shell value. As a result, the attraction coefficient $c_{ij}$ from node $i$ to node $j$ can be formulated, as described in the following equation:

$$c_{ij} = e^{\frac{ks(i)-ks(j)}{ks_{\max}-ks_{\min}}} \tag{19}$$

where *ks*(*i*) and *ks*(*j*) are the k-shell values of nodes *i* and *j* respectively. $ks_{\max}$ and $ks_{\min}$ are the largest and smallest *k*-shell values in the network, $c_{ij}$ represents the gravitational coefficient between node *i* and node *j*, which is equivalent to "*G*" in Eq. (18). When $ks(i) > ks(j)$, $c_{ij} > 1$, similar to how more massive objects exert a greater gravitational pull on neighboring objects, nodes more centrally located in the network have a greater coefficient of attraction. When $ks(i) < ks(j)$, the opposite is true. When $ks(i) = ks(j)$, $c_{ij} = 1$ means that the attraction coefficients between nodes in the same shell are equal. To simplify the analysis, this article focuses only on the first-order neighbors of node *i*.

In a real system, central elements tend to be more stable than non-central elements. This stability arises primarily from the greater interconnectedness of internal components compared to those on the periphery. The $k$-shell value reflects a node's location within a complex network, a higher $k$-shell value suggests that the node occupies a more secure position. This stability enables the node to gain assistance from other nodes situated in the same shell when confronted with external pressures. Nodes with higher *k*-shell values can better withstand external disruptions and exert a greater influence on other nodes. While the *k*-shell value is important, it alone is insufficient to determine the overall significance of a node. Prior studies have indicated that degree centrality represents the direct relationships a node has within a network. In contrast, betweenness centrality evaluates the node's capacity to function as a conduit or intermediary. This measure highlights the significant function of a node in promoting the flow of information by tallying the number of shortest paths that traverse it, linking various nodes throughout the network. This metric evaluates the importance of the position of a node for information propagation by tallying the pathways available for connecting with other nodes. To incorporate these insights, we propose a multi-feature fusion algorithm that combines *k*-shell values with degree centrality and betweenness centrality. Individually, nodes with higher values of these metrics are likely to be more influential, especially when they are in close proximity to other influential nodes. Inspired by principles from gravity, we characterize the combined values of degree centrality, *k*-shell value, and betweenness centrality as the mass of a node. Additionally, the shortest path connecting any two nodes is utilized as a metric to assess their spatial relationship, while the attraction coefficient captures the degree of influence nodes exert on one another. From this framework, the influence of node *i* can be evaluated using the DKBC model outlined below:

$$DKBC(i) = \sum_{d(i,j)\leq R, i\neq j} \frac{c_{ij}(DC(i)+ks(i)+BC(i))(DC(j)+ks(j)+BC(j))}{d^2(i,j)} \tag{20}$$

where $c_{ij}$ is the attraction coefficient, *DC*(*i*) represents the degree centrality of node *i*, *ks*(*i*) represents the *k*-shell value of node *i*, *BC*(*i*) represents the betweenness centrality value of node *i*, and *d*(*i*, *j*) represents the shortest path distance between node *i* and *j*. This model mainly considers three core indexes of node degree centrality, *k*-shell value and betweenness centrality value. For the choice of measurement, we mainly consider these aspects: the degree centrality of a node can be likened to the "weight" of a node, which directly affects the interaction between nodes. Nodes with high degree centrality have more connections in the network and therefore play an important role in information dissemination and influence diffusion. This direct number of connections is an intuitive and critical metric for assessing the force of nodes in a network. The *k*-shell value provides the depth information of the node in the network hierarchy, reflecting the core degree of the node. In the gravity model, this can be regarded as the "potential energy" of the node, that is, the more central the node's position in the network, the higher its importance in the network structure, and the greater the impact on the stability and connectivity of the network. Betweenness centrality measures a node's role as a bridge in a network, that is, the number of times a node appears on the shortest path connecting other pairs of nodes in the network. In the gravity model, the nodes with high betweenness centrality can be regarded as the key "hubs" in the network, which can show good performance in the network with different community structures, and play a crucial role in the information flow and structural integrity of the network. For eigenvector centrality, it measures the influence of a node based on its neighbors. Although this is a global metric, in gravitational models it may not be as straightforward as degree centrality, *k*-shell value, and betweenness centrality as reflecting the "mass" and "position" of a node. It reflects more the indirect influence of nodes than the direct interaction.

Closeness centrality measures the average distance between a node and all other nodes in the network, which reflects the global position of the node to some extent. However, in the gravitational model, we consider the interaction between nodes by dividing by the square of the distance between nodes, which overlaps with the core idea of closeness centrality. Since the gravity model already takes into account the distance between nodes, the introduction of closeness centrality may cause metric redundancy, resulting in increased algorithm complexity without providing additional degree of differentiation. The centrality of degree selection, *k*-shell value and betweenness number is because they can comprehensively evaluate the importance of nodes from different perspectives, covering the number of direct connections of nodes, network core location and global bridge role. This multi-dimensional measurement combination can identify the key nodes in the network more comprehensively, and has certain robustness without the problem of information redundancy. While eigenvector centrality and approach centrality have value in some cases, they may not be as direct and effective as the chosen metric under the framework of gravity-based models.

Nevertheless, we found that the three indicators of DC, *k*-shell and BC do not share the same order of magnitude, making standardization essential. Thus, Eq. (20) can be reformulated as follows:

$$DKBC(i) = \sum_{d(i,j)\leq R, i\neq j} \frac{c_{ij}(\frac{DC(i)}{DC_{\max}} + \frac{ks(i)}{ks_{\max}} + \frac{BC(i)}{BC_{\max}})(\frac{DC(j)}{DC_{\max}} + \frac{ks(j)}{ks_{\max}} + \frac{BC(j)}{BC_{\max}})}{d^2(i,j)} \tag{21}$$

where $DC_{\max}$, $Ks_{\max}$ and $BC_{\max}$ represent the peak values for DC, *k*-shell and BC respectively.

While the normalized *k*-shell index usually surpasses the other two metrics due to its restricted range, it is crucial to lessen the impact of the *k*-shell value. The dominance of lower index values among the majority of nodes stems from the scale-free properties of the network. As a consequence, indices with wider value ranges typically demonstrate a lower median digit-to-maximum value ratio. In the proposed model, both degree centrality and betweenness centrality have significantly broader value ranges compared to the *k*-shell value. For these reasons, we incorporate a discount factor to reduce the influence of the *k*-shell value in the following manner:

$$\alpha = \frac{\max\left\{\frac{DC_{mid}}{DC_{\max}}, \frac{BC_{mid}}{BC_{\max}}\right\}}{\frac{ks_{mid}}{ks_{\max}}} \tag{22}$$

where $DC_{mid}$, $ks_{mid}$ and $BC_{mid}$ represent the median values of DC, *k*-shell and BC respectively. In order to avoid excessive weakening of the function of the *k*-shell value, we take the maximum value of $\left\{\frac{DC_{mid}}{DC_{\max}}, \frac{BC_{mid}}{BC_{\max}}\right\}$

.

Eventually, Eq. (20) can be rephrased as:

$$DKBC(i) = \sum_{d(i,j)\leq R, j\neq i} \frac{c_{ij}(\frac{\varepsilon DC(i)}{DC_{\max}} + \frac{\alpha ks(i)}{ks_{\max}} + \frac{(1-\varepsilon)BC(i)}{BC_{\max}})(\frac{\varepsilon DC(j)}{DC_{\max}} + \frac{\alpha ks(j)}{ks_{\max}} + \frac{(1-\varepsilon)BC(j)}{BC_{\max}})}{d^2(i,j)} \tag{23}$$

where $\varepsilon$ is a tunable parameter, when $\varepsilon = 0$, the proposed model becomes a global information based approach, *k*-shell method and betweenness centrality can show superior performance against sparse networks and networks with strong community structure. When $\varepsilon = 1$, the proposed method is transformed into a hybrid local-global method, which is more efficient when conducting large-scale network operations. The calculation process of DKBC is provided by Algorithm 1.

To illustrate the computational procedure of Algorithm 1, we utilize a basic network depicted in Fig. 1.

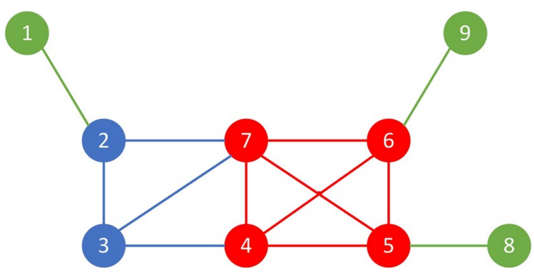


**Fig. 1**. An artificial network, where green indicates the nodes belonging to the 1-shell, blue denotes those in the 2-shell, and red identifies the nodes in the 3-shell.

| Node | DC | *K*-shell | BC |
|---|---|---|---|
| 1 | 0.125 | 1 | 0 |
| 2 | 0.375 | 2 | 0.25 |
| 3 | 0.375 | 2 | 0.0357 |
| 4 | 0.5 | 3 | 0.0714 |
| 5 | 0.5 | 3 | 0.25 |
| 6 | 0.5 | 3 | 0.25 |
| 7 | 0.625 | 3 | 0.393 |
| 8 | 0.125 | 1 | 0 |
| 9 | 0.125 | 1 | 0 |

**Table 2.** The DC value, *k*-shell value and BC value of every node within the artificial network.

| Node | first-order neighbors | second-order neighbors | DKBC |
|---|---|---|---|
| 1 | 2 | 3,7 | 0.4227 |
| 2 | 1,3,7 | 4,5,6 | 4.2136 |
| 3 | 2,4,7 | 1,5,6 | 3.6587 |
| 4 | 3,5,6,7 | 2,8,9 | 10.7669 |
| 5 | 4,6,7,8 | 2,3,9 | 11.5862 |
| 6 | 4,5,7,9 | 2,3,8 | 11.5862 |
| 7 | 2,3,4,5,6 | 1,8,9 | 17.1189 |
| 8 | 5 | 4,6,7 | 0.4187 |
| 9 | 6 | 4,5,7 | 0.4187 |

**Table 3.** The outcome of the DKBC vlaue in the artificial network when $R = 2$.

**Input:** graph:$G < V, E >$, truncation radius: $R$, number of nodes: $N$
**Output:** $Rank[v, DKBC(v)]$
**for** $i \leftarrow 1$ **to** $N$ **do**
  Compute $DC(v_i)$ of node $v_i$;
**end**
**for** $i \leftarrow 1$ **to** $N$ **do**
  Compute $ks(v_i)$ of node $v_i$;
**end**
**for** $i \leftarrow 1$ **to** $N$ **do**
  Compute $BC(v_i)$ of node $v_i$;
**end**
Compute $DC_{\max}$, $Ks_{\max}$, $BC_{\max}$, $DC_{mid}$, $Ks_{mid}$, $BC_{mid}$;
Compute $c_{ij}$ using Eq.19;
Compute $\alpha$ using Eq.22;
**for** $i \leftarrow 1$ **to** $N$ **do**
  Find all neighbors of node $v_i$ within the truncation radius R;
  Compute $DKBC(v_i)$ of node $v_i$ using Eq.23;
**end**
return $Rank[v, DKBC(v)]$;

**Algorithm 1.** Degree-K-shell-betweenness centrality model.

Initially, we calculate the degree centrality, *k*-shell value, and betweenness centrality for the nodes. The results of these computations are presented in Table 2. Here, $\varepsilon$ is set to 0.5.

Secondly, calculate $DC_{\max} = 0.625$, $ks_{\max} = 3$, $BC_{\max} = 0.393$, $DC_{mid} = 0.375$, $ks_{mid} = 2$ and $BC_{mid} = 0.0714$, $ks_{\min} = 1$, and then, obtain $\alpha = 0.9$.

Finally, the DKBC results for $R = 2$ in the network can be found in Table 3. Using node 2 as the reference point, its immediate neighbors include nodes 1, 3, and 7, while the second-level neighbors consist of nodes 4, 5, and 6. The computation yields $c_{ij} = 1.133$, resulting in $DKBC(2) = 4.2136$. In this context, $c_{ij}$ denotes the mean attraction coefficient of node *i* towards all of its adjacent nodes *j*. For node 4 and node 5, we can see from Table 2 that they have the same degree centrality and are in the same k-shell layer. However, we can see from the importance value performance of Table 3 that node 4 is more critical than node 5. This is because node 4 has a higher betweenness centrality than node 5, which means that information between different nodes in the network is more likely to pass through node 4. Similarly, it can be seen from Table 3 that node 7 is the most critical node in the artificial network. Even though it is in the same k-shell layer as nodes 4, 5 and 6, its degree centrality and betweenness centrality are higher than those nodes. Once node 7 is damaged, the communication of a large number of nodes in the network will be interrupted, causing the greatest damage to the network.

### Computational complexity

The computational complexity of the method used in this paper is shown in Table 9. The computing complexity of degree centrality is $O(N+M)$, because it needs to calculate the degree of each node. The computational complexity of *k*-shell algorithm is $O(N+M)$, because it requires several iterations to determine the number of cores for each node, each iteration involves checking the neighbor of the node and its degree. The computational complexity of closeness centrality is $O(NM+N^2\log N)$, due to the need to calculate the distance from each node to all other nodes. The computational complexity of betweenness centrality is $O(NM+N^2\log N$ due to the need to calculate the shortest path between all node pairs. The computational complexity of gravity centrality is $\mathrm{O(N<k>}^R)$ because it computs the R-order neighbors of each node. Most real networks have small-world characteristics, and $R^*=2$ in most cases. Therefore, the computational complexity of gravity centrality generally does not exceed $\mathrm{O(N<k>}^2)$, where $k<<N$. It can be seen that the most computationally complex part of DKBC is the betweenness centrality part, which requires $NM+N^2\log N$ operations. The algorithm complexity is $O(NM+N^2\log N)$.

## Data and experiments

### Datasets

This section presents the analysis of twelve real complex networks spanning various domains to assess the effectiveness of the proposed DKBC. The collection of networks consists of an infrastructure network (Power), a transportation network (USAir), a technology network (Router), two communication networks (Email and EEC), three collaboration networks (Jazz, NS, GrQc), and four social networks (PB, Facebook, WV, PG). The topological characteristics of these networks are summarized in Table 4, which includes metrics such as the number of nodes (*N*), the number of edges (*M*), average degree ($\langle k\rangle$), average path length ($\langle d\rangle$), clustering coefficient (*C*), assortative coefficient (*r*), and the epidemic transmission threshold for the SIR model ($\beta_c$). It is worth mentioning that for the network NS, EEC, PB, Router and GrQc, we set the adjustable parameter $\varepsilon$ to 0.3, because the community structure of the above five networks is more obvious, and giving higher weight to the intermediate centrality can better improve the robustness and performance of the proposed method. In the rest of the network, $\varepsilon$ is set to 0.5.

### Empirical results

Using the described actual network, apply a sophisticated SIR epidemic propagation model and an independent cascade model to evaluate the ranking of key nodes. In the SIR model, a specified transmission probability $\beta$ is used, and in the IC model, the activation probability of each node is determined based on different algorithm metrics. We conducted 1000 independent experiments to ensure the accuracy of the results, and averaged these executions to establish a baseline ranking of node significance. The benchmark algorithms used for comparison include betweenness centrality , closeness centrality, degree centrality, *k*- shell value, eigenvector centrality, gravitational centrality, improved gravity algorithm based on structural hole method (ISM+), M-centrality, and Community Hub Brig (CHB). To determine the accuracy of the algorithm, we tested the diffusion ability of nodes in both the SIR model and the IC model. For the SIR model, we set the propagation probability $\beta=\beta_c$ and the recovery probability $\lambda=1$. We use the top 5 percent of nodes in the corresponding methods as initial

| Networks | N | M | $\langle k\rangle$ | $\langle d\rangle$ | C | r | $\beta_c$ |
|---|---|---|---|---|---|---|---|
| Jazz | 198 | 2742 | 27.6970 | 2.2350 | 0.6334 | 0.0202 | 0.0266 |
| UsAir | 332 | 2126 | 12.8072 | 2.7381 | 0.7494 | -0.2079 | 0.0231 |
| NS | 379 | 914 | 4.8232 | 6.0419 | 0.7981 | -0.0817 | 0.1424 |
| Email | 1133 | 5451 | 9.6222 | 3.6060 | 0.2540 | 0.0782 | 0.0565 |
| PB | 1222 | 16714 | 27.3552 | 2.7374 | 0.3601 | -0.2212 | 0.0125 |
| Facebook | 4039 | 88234 | 43.6911 | 3.6924 | 0.6171 | 0.0635 | 0.0095 |
| Power | 4941 | 6594 | 2.6691 | 18.9892 | 0.1065 | 0.0035 | 0.3483 |
| GrQc | 4158 | 13422 | 6.4559 | 6.0494 | 0.5569 | 0.6392 | 0.0589 |
| EEC | 986 | 16064 | 32.5841 | 2.5869 | 0.4071 | -0.0257 | 0.0136 |
| PG | 6299 | 20776 | 6.5966 | 4.6429 | 0.0109 | 0.0355 | 0.0601 |
| Router | 5022 | 6258 | 2.4922 | 6.4488 | 0.0329 | -0.1384 | 0.0786 |
| WV | 7066 | 100736 | 28.5129 | 3.2475 | 0.2090 | -0.0833 | 0.0069 |

**Table 4.** Topology of twelve real networks.

infected nodes. During the propagation process, infected nodes infect their neighboring nodes with the infection probability $\beta$, and infected nodes become recovery nodes with the recovery probability $\lambda$. Nodes that have already recovered will not be re infected, and the propagation process continues until there are no new recovery nodes in the network. We use Eq. (13) to calculate the number of recovery nodes at time t to determine their

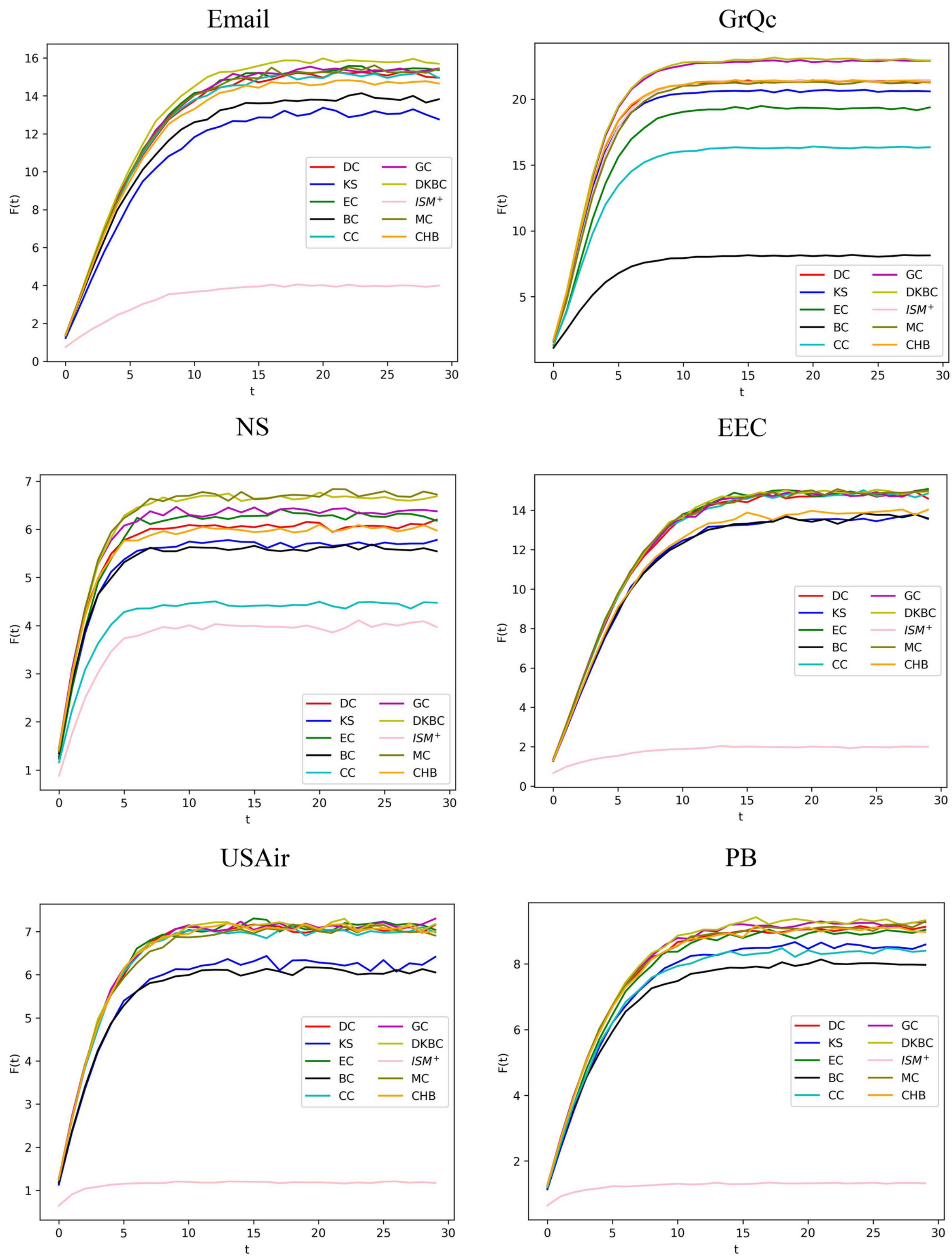


**Fig. 2**. The figure shows a comparison of the average number of recovery nodes produced by various methods based on the SIR Model at time t, using the top five percent of nodes in the network as initial nodes. The propagation rate is set to $\beta = \beta_c$. The obtained curves show that the identified advanced nodes exhibit superior propagation ability in the network.

influence. Figure 2 shows the $F(t)$ value of the node. For the IC model, the higher the value of the metric used, the higher the activation probability of the node, and more nodes can be activated in the same time. We use the top 10 percent of nodes ranked in the corresponding method as the initial activation nodes. The initial activation nodes activate neighboring nodes with activation probability $p_{uv}$. The activation process continues until no

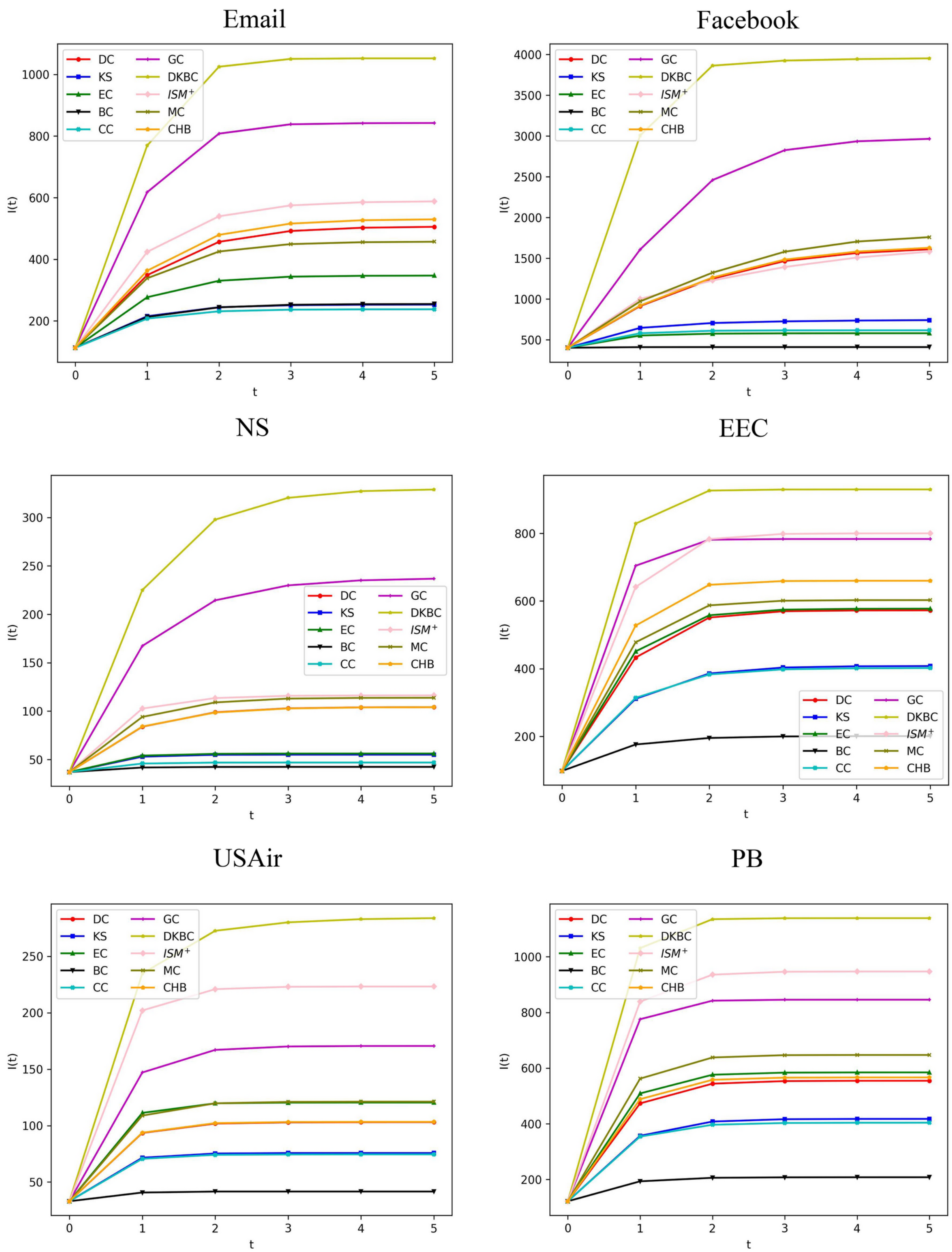


**Fig. 3**. The figure shows a comparison of the average number of active nodes produced by various methods based on the IC model at time t, using the top ten percent of nodes in the network as active nodes. The activation probability is set to $p_{uv}$. The obtained curves show that the nodes with high centrality measure can activate neighbor nodes with greater probability, and exhibit superior propagation ability in the network.

new nodes are activated in the network, ending this operation. The I influence of nodes is shown in Fig. 3. In addition, to determine the effectiveness of the algorithm, we used Kendall's Tau to test the accuracy of each method under different $\beta$ values in the SIR model, and compared the standard rankings generated by the SIR model with the rankings generated by each method. The range of $\tau$ is -1 to 1, where values close to 1 indicate superior performance of the algorithm. The results are shown in Fig. 4, where different $\beta$ values are represented as the ratio between $\beta$ and the propagation threshold $\beta_c$. The five traditional algorithms - betweenness centrality, closeness centrality, degree centrality, eigenvector centrality, and k-shell value - are represented in black, cyan, red, green, and blue, respectively, while the gravity model-based algorithm - GC is represented in magenta,

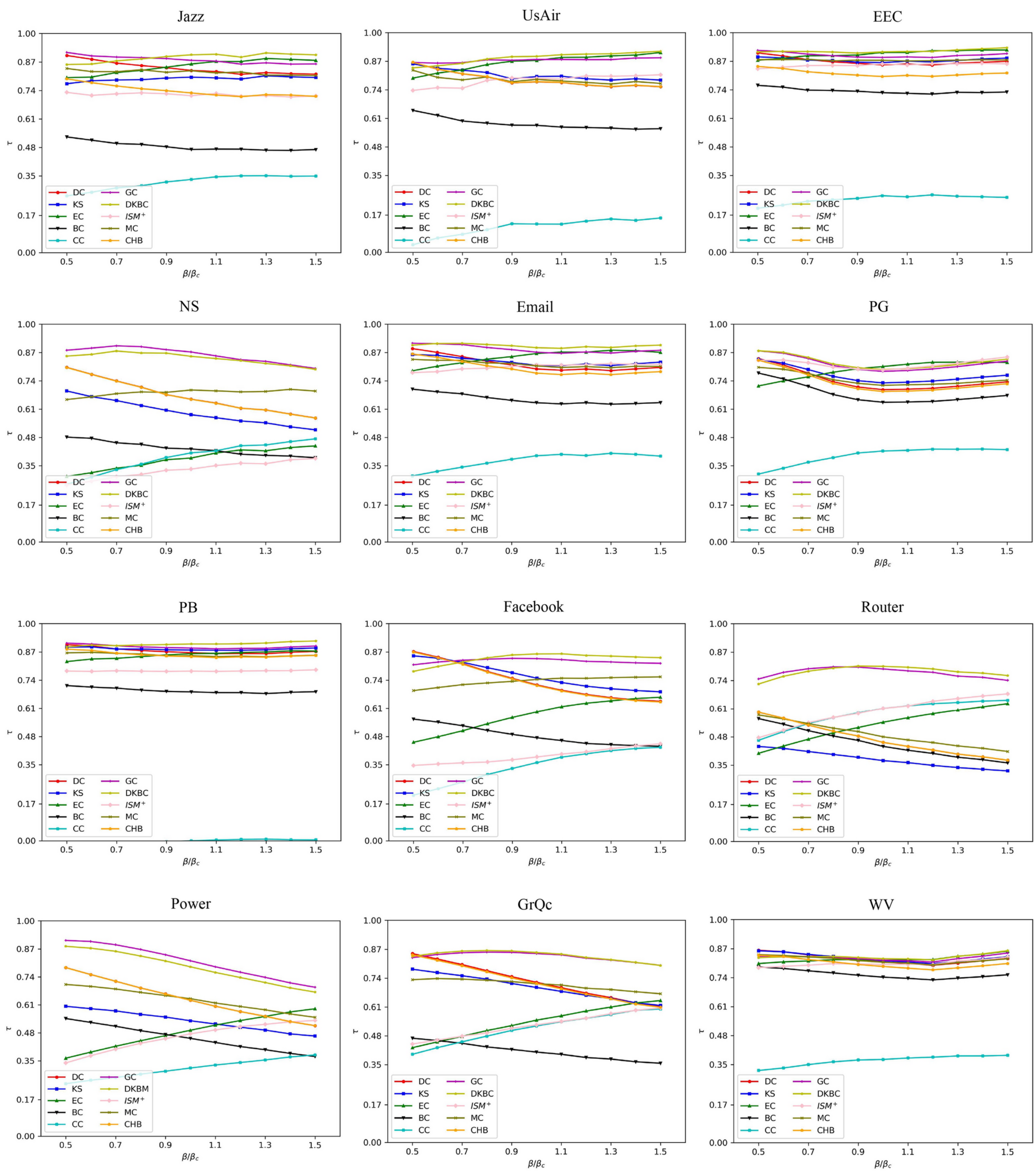


**Fig. 4**. Kendall's Tau was used to test the accuracy of DKBC and other benchmark algorithms under different $\beta$. The top five percent of nodes generated by each method are selected as the initial infected nodes, and the standard ranking list is generated by the SIR model.

ISM+ in pink, and DKBC in gold. M-centrality is represented by olive color, and Community Hub Bridge based on community perception is represented by orange. In the process of calculating CHB, the infomap algorithm is used for community detection.

Figure 2 shows a comparison of $F(t)$ between the various methods. The upward trend of $F(t)$ achieved by DKBC and each benchmark algorithm is largely consistent across all networks. In the Email and GrQc networks, DKBC consistently outperformed its competitors throughout the process. In contrast, in networks such as NS, EEC, and PB, while DKBC is not ahead in the initial stages of the infection process, it is growing faster compared to other methods. In addition, in networks with strong community structures, such as EEC and PB, CHB performs very well. In the later stages of infection, DKBC maintained a high $F(t)$ value across all networks, while the performance of other methods varied across networks. Overall, most of the results described in Fig. 2 show that DKBC performs well in networks of different structures due to its good robustness, indicating that the top-ranked nodes identified by this method have strong propagation ability in the network.

Figure 3 shows the comparison of $I(t)$ among various methods under the independent cascade model. As shown in Fig. 3, the $I(t)$ value of DKBC is higher than that of other algorithms during the entire diffusion process. On Email, USAir and NS, the $I(t)$ obtained by DKBC is approximate to the $I(t)$ obtained by the comparison algorithm at the beginning of the propagation process. However, when diffusion is stable, in EEC, USAir and PB, the advantages of gravity-based approaches are very clear, with DKBC, ISM+, and GC algorithms all outperforming the others on $I(t)$ values. In the network with obvious community structure, CHB shows good competitiveness. It is worth noting that when the transmission process reached a stable state, DKBC acquired the largest number of infections. The experimental results show that the proposed algorithm is an effective network influence node detection algorithm.

Figure 4 shows the Kendall's Tau between the ranking of various methods and the standard ranking generated by the SIR Model. Among them, the standard ranking refers to the node ranking sequence generated by the given initial infected node after running under the SIR Model, which will be used as the "standard answer" to conduct Kendall's Tau correlation analysis with the ranking generated by various methods. The closer the Kendall's Tau is to 1, the better the performance of the method. As shown in Fig. 4, the proposed method is generally superior to other benchmark algorithms in various networks. For example, in Jazz, Email, and EEC networks, DKBC achieved the highest $\tau$ value. Although DKBC may not always maintain the highest ranking among all networks, its overall performance is still relatively high. As the probability of infection increases, especially when it exceeds $\beta_c$, DKBC achieves higher accuracy than classical methods (DC, KS, EC, BC, CC). Compared with gravity model-based methods GC and ISM+, DKBC also demonstrates strong competitiveness on most networks. However, on PG, WV and PB networks, the tau values obtained by DKBC are smaller than those obtained by ISM+, MC, and CHB, but they approach each other as the tau value increases. It is worth noting that, in addition to the three types of networks mentioned above, the performance of KS is significantly better than other classical methods, and even competitive compared to methods based on the law of gravity. This indirectly indicates that methods based on the law of gravity have better stability and their performance will not sharply decline due to differences in networks. Although CHB shows superior $\tau$ values in networks with strong community structures such as Router, PB and GrQc, its performance in networks with weaker community structures is not satisfactory. To present the results more clearly, the average Kendall coefficients are provided in Tables 5 and 6, further supporting our previous observations. Overall, compared with other benchmark algorithms, the proposed DKBC method exhibits better performance.

Upon examination, it was determined that the most effective truncation radius for the majority of networks is around $R = 2$. Therefore, establishing $R = 2$ is sufficient for assessing the efficacy of the DKBC method. In Table 6, the accuracy of DKBC at this setting is compared with that of benchmark algorithms. As can be seen from Tables 5 and 6, gravitational model-based approaches (GC, ISM+ and DKBC) have obvious advantages over other approaches in networks such as Router, NS and Power. This shows that the gravity-model-based approach provides better stability and performance, especially in scenarios with significant network differences, because it preserves the network structure while avoiding rapid degradation in performance. In addition, in

| Networks | DC | KS | EC | BC | CC | GC | ISM+ | MC | CHB | DKBC |
|---|---|---|---|---|---|---|---|---|---|---|
| Jazz | 0.8150 | 0.7638 | 0.8854 | 0.4641 | 0.7008 | 0.8666 | 0.7350 | 0.8421 | 0.7910 | **0.9258** |
| UsAir | 0.7370 | 0.7529 | 0.8946 | 0.5171 | 0.8027 | 0.8875 | 0.2539 | 0.8044 | 0.8438 | **0.9168** |
| NS | 0.5790 | 0.5106 | 0.3660 | 0.3003 | 0.3397 | 0.8372 | 0.4675 | 0.7047 | 0.8048 | **0.8824** |
| Email | 0.7653 | 0.7702 | 0.8832 | 0.6243 | 0.8163 | 0.8697 | 0.4042 | 0.8381 | 0.8684 | **0.9065** |
| PB | 0.8524 | 0.8595 | 0.8738 | 0.6771 | 0.7852 | 0.9030 | 0.1003 | 0.8689 | 0.8794 | **0.9125** |
| Facebook | 0.6798 | 0.7075 | 0.6226 | 0.4529 | 0.3940 | 0.8275 | 0.4285 | 0.7540 | **0.8704** | 0.8695 |
| Power | 0.4264 | 0.3122 | 0.2818 | 0.3254 | 0.3838 | 0.7442 | 0.5364 | 0.7022 | **0.7828** | 0.7683 |
| GrQc | 0.8528 | 0.7803 | 0.6386 | 0.4693 | 0.6046 | 0.8572 | 0.6030 | 0.7343 | 0.8457 | **0.8739** |
| EEC | 0.8063 | 0.8896 | 0.9040 | 0.7618 | 0.8025 | 0.8626 | 0.2621 | 0.8825 | 0.8465 | **0.9210** |
| PG | 0.8405 | 0.8367 | 0.8269 | 0.7752 | 0.8486 | 0.8807 | 0.4286 | 0.8020 | 0.8350 | **0.8910** |
| Router | 0.3139 | 0.1810 | 0.5924 | 0.3096 | 0.6383 | 0.7894 | 0.6480 | 0.5844 | 0.5990 | **0.8078** |
| WV | 0.7619 | 0.7657 | 0.8334 | 0.6978 | 0.8127 | 0.8276 | 0.3905 | 0.8421 | 0.8416 | **0.8569** |

**Table 5**. In the case of $\beta = \beta_c$ in the SIR model, the maximum Kendall's Tau in each network for DKBC and other benchmark algorithms. For each network in the table, the maximum value is highlighted in bold.

| Networks | DC | KS | EC | BC | CC | GC | ISM+ | MC | CHB | DKBC(R = 2) |
|---|---|---|---|---|---|---|---|---|---|---|
| Jazz | 0.8150 | 0.7638 | 0.8854 | 0.4641 | 0.7008 | 0.8666 | 0.7350 | 0.8421 | 0.7910 | **0.9258** |
| UsAir | 0.7370 | 0.7529 | 0.8946 | 0.5171 | 0.2539 | 0.8875 | 0.8027 | 0.8044 | 0.8438 | **0.9168** |
| NS | 0.5790 | 0.5106 | 0.3660 | 0.3003 | 0.3397 | 0.8372 | 0.4675 | 0.7047 | 0.8048 | **0.8796** |
| Email | 0.7653 | 0.7702 | 0.8832 | 0.6243 | 0.4042 | 0.8697 | 0.8163 | 0.8381 | 0.8684 | **0.9065** |
| PB | 0.8524 | 0.8595 | 0.8738 | 0.6771 | 0.7852 | 0.9030 | 0.1003 | 0.8689 | 0.8794 | **0.9125** |
| Facebook | 0.6798 | 0.7075 | 0.6226 | 0.4529 | 0.3940 | 0.8275 | 0.4285 | 0.7540 | **0.8704** | 0.8600 |
| Power | 0.4264 | 0.3122 | 0.2818 | 0.3254 | 0.3838 | 0.7442 | 0.5363 | 0.7022 | **0.7828** | 0.7683 |
| GrQc | 0.8528 | 0.7803 | 0.6386 | 0.4693 | 0.6046 | 0.8572 | 0.6030 | 0.7343 | 0.8457 | **0.8645** |
| EEC | 0.8063 | 0.8896 | 0.9040 | 0.7618 | 0.8025 | 0.8626 | 0.2621 | 0.8825 | 0.8465 | **0.9113** |
| PG | 0.8405 | 0.8367 | 0.8269 | 0.7752 | 0.8486 | 0.8807 | 0.4286 | 0.8020 | 0.8350 | **0.8809** |
| Router | 0.3139 | 0.1810 | 0.5924 | 0.3096 | 0.6383 | 0.7894 | 0.6479 | 0.5844 | 0.5990 | **0.8076** |
| WV | 0.7619 | 0.7657 | 0.8334 | 0.6978 | 0.8127 | 0.8276 | 0.3901 | 0.8421 | 0.8416 | **0.8589** |

**Table 6**. In the case of $\beta = \beta_c$ in the SIR model, let the DKBC algorithm at $R = 2$ and other benchmark algorithms be the optimal Kendall's Tau in each network. For each network in the table, the maximum value is highlighted in bold.

| Networks | DC | KS | EC | BC | CC | GC | ISM+ | MC | CHB | DKBC |
|---|---|---|---|---|---|---|---|---|---|---|
| Jazz | 0.9659 | 0.7944 | 0.9994 | 0.7944 | 0.9885 | 0.9992 | 0.9995 | 0.9982 | 0.9591 | **0.9996** |
| UsAir | 0.8586 | 0.8114 | 0.9951 | 0.8114 | 0.6970 | 0.9950 | 0.9951 | 0.9447 | 0.8586 | **0.9956** |
| NS | 0.7642 | 0.6421 | 0.9953 | 0.6421 | 0.3390 | 0.9933 | 0.9951 | 0.9782 | 0.7631 | **0.9956** |
| Email | 0.8874 | 0.8088 | 0.9999 | 0.8088 | 0.9400 | 0.9998 | **0.9999** | 0.9640 | 0.8790 | 0.9999 |
| PB | 0.9328 | 0.9064 | 0.9993 | 0.9064 | 0.9489 | 0.9993 | 0.9993 | 0.9758 | 0.9257 | **0.9995** |
| Facebook | 0.9739 | 0.9419 | 0.9855 | 0.9419 | 0.9855 | 0.9999 | **0.9999** | 0.9993 | 0.9738 | 0.9999 |
| Power | 0.5927 | 0.2460 | 0.9999 | 0.2460 | 0.8319 | 0.9903 | 0.9968 | 0.8300 | 0.5926 | **0.9999** |
| GrQc | 0.7916 | 0.6925 | 0.9994 | 0.6925 | 0.4850 | 0.9984 | 0.9994 | 0.9422 | 0.7884 | **0.9995** |
| EEC | 0.9571 | 0.9216 | 0.9999 | 0.9216 | 0.9551 | 0.9998 | **0.9999** | 0.9817 | 0.9502 | 0.9999 |
| PG | 0.7638 | 0.5991 | 0.9998 | 0.5991 | 0.8513 | 0.9990 | 0.9997 | 0.8482 | 0.7639 | **0.9998** |
| Router | 0.2886 | 0.0691 | 0.9965 | 0.0691 | 0.3038 | 0.9926 | 0.9948 | 0.3323 | 0.2886 | **0.9969** |
| WV | 0.7761 | 0.7673 | 0.9996 | 0.7673 | 0.7704 | 0.9996 | 0.9996 | 0.8044 | 0.7723 | **0.9996** |

**Table 7**. Monotonicity of DKBC and other benchmark algorithms in twelve networks.

networks with obvious community structure, the measurement of community perceived centrality also shows superior performance. In the gravity-model-based algorithm, DKBC shows the best performance, because it takes into account more reasonable features and parameter Settings. A comparison of twelve real networks shows that DKBC is still highly competitive at different $\beta$values, which indicates that our results are robust to interference to a certain extent.

Compared to benchmark algorithms, DKBC generally performs best when $R = 2$, achieving nearly optimal results across the twelve real networks. Nonetheless, in situations where the network's average distance is notably high, exemplified by a value of 18.9892 for the Power network, configuring $R = 2$ can substantially affect the effectiveness of DKBC. This influence arises from the fact that the ideal truncation radius generally demonstrates a linear relationship with the average distance. Fortunately, the majority of real-world networks exhibit small-world properties, meaning that the optimal truncation radius $R^*$ is typically small and does not significantly affect the performance of DKBC in most scenarios.

Furthermore, monotonicity is used to further evaluate the differentiability of different algorithms. Table 7 shows the monotonicity values of DKBC on twelve networks and nine benchmark algorithms, with the maximum monotonicity value of the algorithm in each network highlighted in bold. Experimental results show that the proposed algorithm has the best monotonicity on most networks. In addition, the monotonicity values of GC, ISM+, MC, and CHB are also close to DKBC on most datasets. We noticed that ISM+ peaked in a small subset of networks, but its overall performance was inferior to DKBC. In conclusion, the proposed DKBC algorithm performs best in terms of monotonicity, and it can assign a unique influence value to each node in the network to a certain extent.

Finally, we compare DKBC implemented with Eqs. (20), (21) and (23) to highlight the importance of normalization and the introduction of discount factors. As shown in Table 8, DKBC is gradually improved in the standardization process, which emphasizes the importance of standardization and the effectiveness of the standardization strategy adopted.

| Networks | DKBC (Eq.20) | DKBC (Eq.21) | DKBC (Eq.23) |
|---|---|---|---|
| Jazz | 0.8888 | 0.9212 | **0.9258** |
| UsAir | 0.8946 | 0.9060 | **0.9168** |
| NS | 0.8428 | 0.8710 | **0.8824** |
| Email | 0.8782 | 0.8986 | **0.9065** |
| PB | 0.9047 | 0.9110 | **0.9125** |
| Facebook | 0.8381 | 0.8547 | **0.8695** |
| Power | 0.7557 | 0.7569 | **0.7683** |
| GrQc | 0.8569 | 0.8643 | **0.8739** |
| EEC | 0.9154 | 0.9189 | **0.9210** |
| PG | 0.8892 | 0.8902 | **0.8910** |
| Router | 0.7992 | 0.7983 | **0.8386** |
| WV | 0.8299 | 0.8341 | **0.8369** |

**Table 8**. The performance of the DKBC algorithm, assessed through Eqs. (20), (21), and (23), is analyzed using Kendall's Tau with $\beta = \beta_c$ in SIR model. Parameters are meticulously adjusted to achieve their best values for optimizing $\tau$. Each network's most effective algorithm is distinguished by being presented in bold.

| Methods | Complexity |
|---|---|
| DC | $O(N+M)$ |
| KS | $O(N+M)$ |
| EC | $O(N+M)$ |
| BC | $O(NM+N^2 \log N)$ |
| CC | $O(NM+N^2 \log N)$ |
| GC | $\mathrm{O(N<k>^{\mathit{R}})}$ |
| ISM$^+$ | $\mathrm{O(N<k>^2)}$ |
| MC | $O(N+M)$ |
| CHB | $\mathrm{O(N^2C)}$ |
| DKBC | $O(NM+N^2 \log N)$ |

**Table 9**. The computational complexity of DKBC and each benchmark algorithm.

## Conclusions

In this study, we present the DKBC model, an innovative gravity model that successfully incorporates the features of multiple nodes. The framework takes into account factors such as the number of adjacent nodes, the distance between nodes, the location of nodes in the network, the specifics of the connections between nodes, and the ability of nodes to act as intermediaries. In addition, we implemented a standardization strategy and introduced discount factors and tunable parameters to solve the problem of multiple indicators with different orders of magnitude and the ability to handle networks with different structures, a process called standardization. The experimental results demonstrate the importance of standardization and the effectiveness of this method. In order to verify the performance of DKBC, the experiment was conducted in twelve real networks, and the algorithm was compared with several mainstream algorithms such as ISM+, MC, GC and CHB, and compared with five existing classical algorithms. Through the empirical analysis of the diffusion dynamics of the SIR Model and the IC model, the proposed DKBC algorithm achieves the maximum extended influence when the diffusion process terminates. The experimental results show that the DKBC algorithm is superior to the nine benchmark algorithms in accuracy and differentiation. In addition, DKBC scored highest in the uniqueness of rankings, achieving almost unique rankings across twelve datasets, and even achieving a monotonicity score of 0.9999. However, it is undeniable that the computational complexity of this method is higher in large networks due to the existence of betweenness centrality. Although the efficiency can be improved by setting the adjustable parameter $\varepsilon$ to 1 and ignoring the calculation of the betweenness centrality, the performance may not reach the original effect, which is also the limitation of the method. Going forward, we will continue to explore this area, applying it to dynamic networks, heterogeneous networks and multilayer networks. Investigating how to adapt these types of networks in future research.

## Data availability

The datasets of complex networks, including Jazz, UsAir, NS, Email, PB, Facebook, Power, GrQc, EEC, PG, Rputer, and WV are available for download from the web of Github at https://github.com/MLIF/Network-Data.

Received: 16 August 2024; Accepted: 12 March 2025

Published online: 03 April 2025

## References


1. Gallos, L. K., Sigman, M. & Makse, H. A. The conundrum of functional brain networks: Small-world efficiency or fractal modularity. *Front. Physiol.* **3**, 123 (2012).
2. Wei, B., Xiao, F. & Shi, Y. Synchronization in Kuramoto oscillator networks with sampled-data updating law. *IEEE Trans. Cybern.* **50**, 2380–2388 (2019).
3. Wei, X., Zhao, J., Liu, S. & Wang, Y. Identifying influential spreaders in complex networks for disease spread and control. *Sci. Rep.* **12**, 5550 (2022).
4. Shang, Q., Zhang, B., Li, H. & Deng, Y. Identifying influential nodes: A new method based on network efficiency of edge weight updating. *Chaos Interdiscipl. J. Nonlinear Sci.* **31** (2021).
5. Zhao, J., Song, Y., Liu, F. & Deng, Y. The identification of influential nodes based on structure similarity. *Connect. Sci.* **33**, 201–218 (2021).
6. Newman, M. E. A measure of betweenness centrality based on random walks. *Soc. Netw.* **27**, 39–54 (2005).
7. Sabidussi, G. The centrality index of a graph. *Psychometrika* **31**, 581–603 (1966).
8. Freeman, L. C. et al. Centrality in social networks: Conceptual clarification. *Soc. Netw. Crit. Concepts Sociol. Londres Routledge* **1**, 238–263 (2002).
9. Bonacich, P. & Lloyd, P. Eigenvector-like measures of centrality for asymmetric relations. *Soc. Netw.* **23**, 191–201 (2001).
10. Kitsak, M. et al. Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893 (2010).
11. Chen, D., Lü, L., Shang, M.-S., Zhang, Y.-C. & Zhou, T. Identifying influential nodes in complex networks. *Phys. A Stat. Mech. Appl.* **391**, 1777–1787 (2012).
12. Lü, L. et al. Vital nodes identification in complex networks. *Phys. Rep.* **650**, 1–63 (2016).
13. Kumar, S. & Panda, B. Identifying influential nodes in social networks: Neighborhood coreness based voting approach. *Phys. A Stat. Mech. Appl.* **553**, 124215 (2020).
14. Alshahrani, M., Fuxi, Z., Sameh, A., Mekouar, S. & Huang, S. Efficient algorithms based on centrality measures for identification of top-k influential users in social networks. *Inf. Sci.* **527**, 88–107 (2020).
15. Wen, T., Pelusi, D. & Deng, Y. Vital spreaders identification in complex networks with multi-local dimension. *Knowl.-Based Syst.* **195**, 105717 (2020).
16. Xiao, L., Wang, S. & Mei, G. Efficient parallel algorithm for detecting influential nodes in large biological networks on the graphics processing unit. *Future Gener. Comput. Syst.* **106**, 1–13 (2020).
17. Zareie, A., Sheikhahmadi, A. & Jalili, M. Influential node ranking in social networks based on neighborhood diversity. *Future Gener. Comput. Syst.* **94**, 120–129 (2019).
18. Zareie, A., Sheikhahmadi, A. & Jalili, M. Identification of influential users in social networks based on users' interest. *Inf. Sci.* **493**, 217–231 (2019).
19. Sheikhahmadi, A., Nematbakhsh, M. A. & Zareie, A. Identification of influential users by neighbors in online social networks. *Phys. A Stat. Mech. Appl.* **486**, 517–534 (2017).
20. Hu, Z.-L., Liu, J.-G., Yang, G.-Y. & Ren, Z.-M. Effects of the distance among multiple spreaders on the spreading. *Europhys. Lett.* **106**, 18002 (2014).
21. Bian, T. & Deng, Y. Identifying influential nodes in complex networks: A node information dimension approach. *Chaos Interdiscip. J. Nonlinear Sci.* **28** (2018).
22. Ren, Z.-M., Liu, J.-G., Shao, F., Hu, Z.-L. & Guo, Q. Analysis of the spreading influence of the nodes with minimum k-shell value in complex networks. *Acta Phys. Sin.* **62**, 108902 (2013).
23. Zhang, J.-X., Chen, D.-B., Dong, Q. & Zhao, Z.-D. Identifying a set of influential spreaders in complex networks. *Sci. Rep.* **6**, 27823 (2016).
24. Guo, L., Lin, J.-H., Guo, Q. & Liu, J.-G. Identifying multiple influential spreaders in term of the distance-based coloring. *Phys. Lett. A* **380**, 837–842 (2016).
25. Sheikhahmadi, A. & Nematbakhsh, M. A. Identification of multi-spreader users in social networks for viral marketing. *J. Inf. Sci.* **43**, 412–423 (2017).
26. Liu, F., Wang, Z. & Deng, Y. GMM: A generalized mechanics model for identifying the importance of nodes in complex networks. *Knowl.-Based Syst.* **193**, 105464 (2020).
27. Yu, E.-Y., Wang, Y.-P., Fu, Y., Chen, D.-B. & Xie, M. Identifying critical nodes in complex networks via graph convolutional networks. *Knowl.-Based Syst.* **198**, 105893 (2020).
28. Berahmand, K., Bouyer, A. & Samadi, N. A new local and multidimensional ranking measure to detect spreaders in social networks. *Computing* **101**, 1711–1733 (2019).
29. Samadi, N. & Bouyer, A. Identifying influential spreaders based on edge ratio and neighborhood diversity measures in complex networks. *Computing* **101**, 1147–1175 (2019).
30. Zhong, S., Zhang, H. & Deng, Y. Identification of influential nodes in complex networks: A local degree dimension approach. *Inf. Sci.* **610**, 994–1009 (2022).
31. Ullah, A. et al. Identifying vital nodes from local and global perspectives in complex networks. *Expert Syst. Appl.* **186**, 115778 (2021).
32. Yi-Run, R., Song-Yang, L., Jun, T., Liang, B. & Yan-Ming, G. Node importance ranking method in complex network based on gravity method. *Acta Phys. Sin.* **71** (2022).
33. Rajeh, S., Savonnet, M., Leclercq, E. & Cherifi, H. Comparative evaluation of community-aware centrality measures. *Qual. Quant.* **57**, 1273–1302 (2023).
34. Rajeh, S., Savonnet, M., Leclercq, E. & Cherifi, H. Characterizing the interactions between classical and community-aware centrality measures in complex networks. *Sci. Rep.* **11**, 10088 (2021).
35. Ma, L.-L., Ma, C., Zhang, H.-F. & Wang, B.-H. Identifying influential spreaders in complex networks based on gravity formula. *Phys. A Stat. Mech. Appl.* **451**, 205–212 (2016).
36. Liu, F., Wang, Z. & Deng, Y. GMM: A generalized mechanics model for identifying the importance of nodes in complex networks. *Knowl.-Based Syst.* **193**, 105464 (2020).
37. Shang, Q., Deng, Y. & Cheong, K. H. Identifying influential nodes in complex networks: Effective distance gravity model. *Inf. Sci.* **577**, 162–179 (2021).
38. Yang, X. & Xiao, F. An improved gravity model to identify influential nodes in complex networks based on k-shell method. *Knowl.-Based Syst.* **227**, 107198 (2021).
39. Xu, G. & Dong, C. Cagm: A communicability-based adaptive gravity model for influential nodes identification in complex networks. *Expert Syst. Appl.* **235**, 121154 (2024).
40. Ibnoulouafi, A., El Haziti, M. & Cherifi, H. M-centrality: Identifying key nodes based on global position and local degree variation. *J. Stat. Mech. Theory Exp.* **2018**, 073407 (2018).
41. Ghalmane, Z., Hassouni, M. E. & Cherifi, H. Immunization of networks with non-overlapping community structure. *Soc. Netw. Anal. Min.* **9**, 1–22 (2019).
42. Rajeh, S. & Cherifi, H. On the role of diffusion dynamics on community-aware centrality measures. *Plos One* **19**, e0306561 (2024).

## Acknowledgements
This work is supported by the National Natural Science Foundation of China under grant No. 62273294 and 62103354, the Science and Technology Project of Hebei Education Department under grant No. ZD2022104, the Natural Science Foundation of Hebei under grant no. F2022203081, and the Funding Project for the Introduced Overseas Students of Hebei Province under grant No. C20220337.

## Author contributions
Shaobao Li: Conceptualization, Investigation, Writing - Original Draft, Supervision, Funding acquisition. Yiran Quan: Methodology, Investigation, Validation, Writing - Original Draft. Xiaoyuan Luo: Writing - Review & Editing, Supervision. Juan Wang: Writing - Review & Editing, Project administration.

## Declarations

## Competing interests
The authors declare no competing interests.

## Additional information
**Correspondence** and requests for materials should be addressed to J.W.