Omega 122 (2024) 102945

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

ERIUE: Evidential reasoning-based influential users evaluation in social networks☆

Tao Wen, Yu-wang Chen *, Tahir abbas Syed, Ting Wu

Alliance Manchester Business School, The University of Manchester, Manchester M15 6PB, UK

ARTICLE INFO

Keywords:
Social network
Evidential reasoning
Influential users
Decision-making

ABSTRACT

Social media users are playing an increasingly important role in disseminating information, but their ability to diffuse information may vary significantly. Therefore, evaluating the influential ability of users has become crucial to promote or curb the dissemination of specific information. Existing centrality measures have produced varying results in identifying the most influential users, but it remains a challenge to identify the most influential users in a multifaceted and consistent way in social networks, especially when only a limited number of users can be nominated. To fill this gap, this work developed an evidential reasoning-based influential users evaluation (ERIUE) model that considers multiple sources of structural information in networks. Our proposed model collates information about users' influential ability from multiple forms of centrality measures and maps their scores to different grades in an informative belief distribution. To determine the weight of each centrality, three types of information are considered: conflict of belief distributions, similarity of probability sets, and overlap of evaluations. The information is aggregated using the recursive evidential reasoning approach based on a formulated criterion hierarchy, thereby determining the influential ability of users. The applicability of our proposed model is demonstrated by comparing it with existing measures in three real-world social networks. Our proposed model is also applicable to relevant problems beyond identifying influential users, including preventing epidemic spread, cascade failure, and misinformation dissemination in social networks.

1. Introduction

Social media has transformed the way people communicate and share information, enabling them to connect with friends and even strangers regardless of time and space. With a large number of internet users participating in social media, massive amounts of information are generated and disseminated. To understand the role of individuals and information dissemination in social media, social networks, in which users and their connections are represented by nodes and edges respectively, have been used to describe the interaction between users [1,2]. Some important micro and macro properties have been explored to increase the understanding of the networks, for example, predicting potential connections between users to help them meet new people [3,4], constructing recommender systems to recommend suitable products and links to users [5,6], studying the way to reach consensus in complex interactions to guide large-scale group decision-making [7,8], exploring properties of the community to analyse user portraits [9,10], and assessing the risk of transportation or financial networks to mitigate losses [11,12]. In these studies, evaluating the influential ability of users in social networks has gradually become important [13–15] due to their impact on the dissemination of information. In social networks, users' influential abilities differ due to their varying connectivity. For instance, users with more friends are known to be influential, represented by degree centrality in the network, while users on the bridging edge (betweenness centrality [16]) are influential due to their mediating effect on different types of users. Highly influential users can disseminate information more widely and rapidly [17]. Therefore, identifying such users in advance can help platform operators promote or curb the spread of specific information, for example, misinformation about vaccines during COVID-19 pandemic [18].

There has been immense interest in the identification of influential users (also known as influence maximization [19,20]) on social media. To address this issue, various centrality measures have been proposed. Classic centrality measures include betweenness centrality, closeness centrality, eigenvector centrality, and H-index methods, which primarily consider factors such as the number of neighbour nodes and the shortest path between nodes [13,21]. However, these existing centralities suffer from certain limitations, including inadequate consideration of information, limited applicability in specific scenarios, or

☆ This manuscript was processed by Associate Editor Prof. Benjamin Lev.

* Correspondence to: Alliance Manchester Business School, The University of Manchester, Manchester M15 6PB, UK.
E-mail address: Yu-wang.Chen@manchester.ac.uk (Y.-w. Chen).

https://doi.org/10.1016/j.omega.2023.102945
Received 1 May 2023; Accepted 10 August 2023
Available online 19 August 2023
0305-0483/

high time complexity. For instance, degree centrality only takes into account the local structural information surrounding each node, which is insufficient to capture the overall information propagation ability within the network. Eigenvector centrality and closeness centrality, on the other hand, are only suitable for symmetric networks and connected networks, respectively. Meanwhile, betweenness centrality requires considering all paths between every pair of nodes, resulting in high computational complexity, especially in large-scale networks. In recent years, researchers from various fields have explored this issue and proposed alternative approaches. For example, Google introduced the PageRank algorithm, which utilizes random walks to rank search engine results [22]. The Hyperlink-induced topic search (HITs) algorithm was also developed to assess the authority and hub status of websites in the World Wide Web [23]. Physicists have applied classical physical theories to tackle this problem as well. For example, the k-shell decomposition analysis revealed that the most influential users tend to be located at the core of social networks, rather than the periphery [15]. Dimensionality has also been considered to explore the impact of fractal property on node influential ability [24,25]. However, most centrality measures focus on only one aspect of the network's topology, leading to a biased evaluation of nodes' influential ability due to the limited information considered. Therefore, it is crucial to incorporate multiple sources of structural characteristics to accurately identify influential users in social networks.

Multiple criteria decision-making (MCDM) models are valuable tools for integrating various sources of structural information in network analysis. This complex process involves considering multiple criteria and relies on the cognitive insights and knowledge of different stakeholders, such as individuals, institutions, or countries, to identify and select alternatives. Initially, decision-making approaches were based on economic models [26]. However, with technological advancements, several MCDM models have been proposed to guide decision-making processes, including fuzzy set theory, Analytic Hierarchy Process (AHP), and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). Recently, novel MCDM models have emerged to address different application scenarios. For example, there are models that account for uncertain pairwise comparisons between alternatives [27], facilitate the selection of preferred alternatives in choice problems [28], and enable data-driven decision-making by considering interrelationships between criteria [29]. These models have been applied to address issues in different fields, for example, they have been used to assess the contribution of policies to the set of objectives of the European Union marine litter strategy [30], assist patients in making medical choices on online medical platforms [31], understand consumer preferences to promote the adoption of alternative fuel vehicles [32], and evaluate the implementation of circular economy practices in hotels to suggest improvements for achieving higher circularity levels [33]. One notable MCDM approach is Dempster–Shafer evidence theory [34], which provides a framework for reasoning with uncertainty using belief functions rather than probability distributions. In this framework, probabilities are assigned to subsets representing all possible combinations of propositions in the powerset of the frame of discernment. The evidential reasoning (ER) algorithm [35] was proposed to rank alternatives by considering both quantitative and qualitative attributes with uncertainty. To handle highly or completely conflicting pieces of evidence, the ER algorithm introduced the weight and reliability [36], which enhance Dempster's rule and facilitate its application. Based on a new likelihood analysis, the evidence can be also obtained from imperfect data [37], enabling likelihood inference within the ER framework. This likelihood inference process can degenerate into Bayesian inference under some given conditions. Linguistic belief structures have also been integrated into the ER framework [38], where the weight of experts is determined by their hesitancy. These advanced ER-based models have been applied to railway track maintenance management [37] and lung cancer diagnosis [38]. Given the complex nature of network structures, belief functions have been modelled using network graphs [39]. This approach addresses typical challenges such as conflicting evidence and evidence clustering. The ER framework, with its ability to consider uncertainty and aggregate information from different agents, has been applied to explore vaccination decision-making under social influence in social networks [40]. Evidence theory has also been utilized to identify key players in social networks [41]. As a result, a close relationship has been established between evidence-based methods and social network analysis.

In this paper, we present the Evidential Reasoning-based Influential Users Evaluation (ERIUE) model, a comprehensive approach for identifying influential nodes in social networks by considering multiple sources of structural information. The model utilizes a two-level criterion hierarchy and collects structural information around nodes from three types of criterion: neighbourhood-based centrality, path-based centrality, and iterative refinement centrality. Each type includes three classic centralities as sub-criteria, and the scores obtained by each centrality are normalized and mapped to a belief distribution that indicates a node's influential ability. The weight of each criterion is determined using three aspects of information, including the conflict in belief distributions describing influential ability, the similarity of scores expressed as probability sets, and the degree of overlap of scores. The overall belief distribution can be aggregated by the recursive ER approach within the formulated two-level criterion hierarchy, providing a measure of the influential ability of nodes. This approach allows for the combination of structural properties at different scales with varying weights, enabling a comprehensive evaluation of node importance. The applicability of our proposed model is demonstrated in three real-world social networks by comparing it with existing measures. The correlation between the results obtained by each centrality and the benchmark method (Susceptible–Infected–Recovered model) is measured by Kendall's Tau correlation coefficient and the Jaccard similarity coefficient, thereby demonstrating the effectiveness of our proposed model. Our experiments show that the ERIUE model considers sufficient structural information, resulting in a ranking list that exhibits a higher correlation with the benchmark method under different parameter settings in the three networks. The main contributions of this work are summarized as follows,

1. A novel evaluation model is developed to identify influential nodes in social networks by leveraging multiple criteria decision-making, and several existing centrality measures that consider different types of structural information are aggregated with weight parameters.
2. The performance of each criterion is comprehensively evaluated from three perspectives, taking into account the conflict in belief distributions describing influential ability, the similarity of score lists, and the degree of overlap of scores. This comprehensive evaluation ensures a rational and reliable decision-making process.
3. Three real-world social networks are applied to demonstrate the performance and applicability of our proposed model. The experiment results indicate that our proposed method can outperform classical centrality measures in most parameter settings and yield a ranking list that closely matches the list obtained by the benchmark method.

The remainder of this paper is organized as follows. Some existing influential ability identification models in social networks are introduced in Section 2. Section 3 develops the proposed influential ability identification model in four parts. The numerical simulations in three real-world networks are given in Section 4. Finally, the conclusions are discussed in Section 5.

2. Influential ability identification methods in social networks

In the social network, the ability of individuals to diffuse information or spread a disease is influenced by the topological structure of the network, particularly the local structure around the node. Therefore, various influential ability identification models are introduced in this section, including classical centrality measures as well as methods from different disciplines. In addition, we introduce different information diffusion models that can be used as benchmark methods and explain why they cannot often be used in large-scale networks as a method for identifying vital nodes.

2.1. Classical centrality measures

The influential ability of users is strongly influenced by their positions within the social network and their connection with friends their connections with other individuals, including their friends and their friends' connections with others. Therefore, the structural characteristics of the social network play a crucial role in determining the influential ability of users. To capture these structural properties, centrality measures have proven to be effective tools, and as a result, several centrality measures have been utilized to identify and rank the influential ability of users in social networks [13].

The first type of centrality measures mainly considers the information from the neighbourhood. For a given social network $\mathcal{G}(\mathcal{N},\mathcal{E})$, the set of nodes and edges are represented by $\mathcal{N}=\{1,2,\dots,i,\dots,|\mathcal{N}|\}$ and $\mathcal{E}=\{(i,j)\ :\ i,j\in\mathcal{N}\}$, respectively, where the number of nodes and edges are $|\mathcal{N}|$ and $|\mathcal{E}|$, respectively. The network topological structure is indicated by the adjacency matrix $\mathbf{A}$, whose element a_{ij} is defined as,

$$a_{ij}=\begin{cases}1, & (i,j)\in\mathcal{E}\\ 0, & (i,j)\notin\mathcal{E}\end{cases}. \tag{1}$$

Usually, the network is an undirected network ($a_{ij}=a_{ji}$), but it becomes directed if $a_{ij}\neq a_{ji}$ (asymmetric $\mathbf{A}$). In an undirected network, the neighbourhood set of node i is represented by $\mathcal{N}_i=\{j\ :\ (i,j)\in\mathcal{E}\}$. The simplest centrality that considers the information from the neighbourhood is the degree centrality (DC),

$$k_i=|\mathcal{N}_i|=\sum_{j\in\mathcal{N}_i}a_{ij}, \tag{2}$$

which describes the number of nodes in the neighbourhood set. The more neighbours node i has, the more influential it is. It is also applied to find an influential cohesive cluster of nodes rather than a single node [42]. Gradually, researchers found that only considering the number of neighbours is not sufficient to reflect the local structure and evaluate the influence of nodes. Hence, two types of information are further considered. The first type is to consider higher-order neighbours' information, and the LocalRank [43] was developed to consider the information up to fourth-order neighbours,

$$C_{LR}(i)=\sum_{j\in\mathcal{N}_i}\sum_{k\in\mathcal{N}_j}R(k), \tag{3}$$

where $R(k)$ is the number of the first- and second-order nearest neighbours of node k. Another type is to consider the location of nodes because it has been found that a node located in the core part of the network is more influential than that located in the periphery [15]. Hence, the coreness [15] was developed to evaluate the spreading influence of nodes by the k-shell decomposition. Isolated nodes are removed before the decomposition, leading their coreness to be $c_i=0$. In the k-shell (KS) decomposition, nodes whose residual degrees $k_i\le 1$ are continuously removed until all remaining nodes' residual degrees $k_i>1$. The removed nodes are in the 1-shell and their coreness is $c_i=1$. This process continues until all nodes are removed from the network, and all nodes are assigned with different c_i according to their shell layer. Due to its high computational complexity and the requirement of global topological structure, the coreness is difficult to be obtained in large-scale dynamic networks. The H-index (HI) [44] has been developed, which considers the number and degree of neighbours. It is expressed as the largest value of h_i such that node i has at least h_i neighbours with a degree of at least h_i, which is the same as the H-index describing the scientific research output of scholars [45].

Another type of centrality considers the information from the paths, where the shortest distance between nodes is the basis. The shortest path d_{ij} between nodes (in the shortest distance matrix $\mathbf{D}$) can be obtained by Floyd or Dijkstra's algorithms to describe the pairwise relationship,

$$d_{ij}=a_{il_1}+a_{l_1l_2}+\cdots+a_{l_{n-1}l_n}+a_{l_nj}, \tag{4}$$

where $l_1,l_2,\dots,l_n$ are IDs of nodes on the shortest path between nodes i and j. The simplest centrality based on the shortest distance is the eccentricity centrality [46]. It represents the maximum value of the distance of the shortest path from the central node to the others,

$$C_{ECC}(i)=\max_{j\neq i}d_{ij}. \tag{5}$$

By summarizing the shortest distance from the central node to the rest of nodes, the closeness centrality (CC) [47] was developed,

$$C_C(i)=\frac{|\mathcal{N}|-1}{\sum_{j\neq i}d_{ij}}. \tag{6}$$

This centrality reflects the efficiency of information exchange between the central node and the others, which has been used to describe the average efficiency of nodes [48]. The power of a node in controlling the information flow in networks can be also indicated by the betweenness centrality (BC) [16],

$$C_B(i)=\sum_{i\neq j,i\neq k,j\neq k}\frac{g_{jk}^i}{g_{jk}}, \tag{7}$$

where g_{jk} is the number of the shortest paths between node j and k, and g_{jk}^i is the number of shortest paths between node j and k that pass through node i. Several variants of BC have been developed to consider the difference between paths, such as communicability BC [49] and random-walk BC [50]. By considering the shortest distance and degree of nodes, a gravity-based centrality (GC) [51] was developed inspired by the gravity law,

$$C_G(i)=\sum_{i\neq j}\frac{k_ik_j}{d_{ij}}, \tag{8}$$

which also needs the global structural information of the network. There are still numerous centrality measures to evaluate the influential ability of nodes based on the information from paths, such as the information index, Katz index, and subgraph centrality. Interested readers can refer to [13] for details.

Due to the mutual enhancement effect [52] where the influential ability of a node depends on the number and influence of neighbours, another centrality measure (iterative refinement centrality) determines the influential ability by iteratively calculating the influential ability of each node and its neighbours in the entire network. The eigenvector centrality (EC) [53] that considers both the quantity and quality of neighbours is obtained by,

$$C_E(i)=c\sum_{j\in\mathcal{N}}a_{ij}C_E(j), \tag{9}$$

where c is a proportionality constant related to the largest eigenvalue of $\mathbf{A}$, and the initial value of C_E for all nodes equals to 1. There are several variants to further distinguish the nodes under common conditions, such as nonbacktracking centrality [54] and alpha centrality [55]. The most famous variant is the PageRank (PR) algorithm [22] which was developed by Google to rank websites in the search engine. The PR value is iteratively obtained by,

$$PR^{(t)}(i)=s\sum_{j\in\mathcal{N}}a_{ji}\frac{PR^{(t-1)}(j)}{k_j^{out}}+(1-s)\frac{1}{n}. \tag{10}$$

where k_j^{out} is the out-degree of node j, and s is the random jumping factor used to avoid the system from not converging (due to the existence of the dangling node). Another iterative refinement centrality is called HITs algorithm [23]. In this centrality, node i has the role of authority $a_i(t)$ and hub $h_i(t)$ where authoritative nodes are reliable and hub nodes usually connect to many related authorities. The values can be obtained by,

$$\begin{aligned} a_i'(t) &= \sum_{j\in\mathcal{N}} a_{ji} h_j(t-1), \\ h_i'(t) &= \sum_{j\in\mathcal{N}} a_{ij} a_j'(t), \end{aligned} \tag{11}$$

after normalization. Most of the centrality measures mentioned take into account a (limited) structural information and thus are not effective in identifying influential nodes.

2.2. Recent influential ability identification models

Since the network is a complicated structure, researchers in different fields developed several models to identify key nodes in networks. They found that the optimal (minimal) set of structural nodes, much smaller than the network size, can efficiently spread information or prevent the diffusion of epidemics throughout the whole network. As a result, heuristic strategies and evolution optimization methods have been applied to find the optimal set of nodes, incorporating different objective functions, mutation schemes, and optimization algorithms, such as the moth-flame optimization method [56] and the percolation-based framework [57]. However, these approaches often lack a global function of influence and are tailored to specific problems, which limits their performance and applicability in general scenarios [58,59].

With the advent of artificial intelligence models, several machine learning models have also been applied to find the key players in large-scale networks. A typical model is the deep reinforcement learning framework FINDER [58], which has shown promising performance in various application scenarios despite being trained on small networks generated by toy models. This opens up new avenues for leveraging artificial intelligence to gain insights into the principles of network structure. However, the mechanisms underlying deep learning frameworks are complex and require further exploration, as their interpretability is a challenge.

From the perspective of physics, a general dismantling framework has been proposed to dismantle the network into more isolated components by removing a minimal number of nodes [60]. This process is akin to identifying and removing a set of important nodes one by one. This framework combines the spectral analysis of the Laplace matrix and fine-tuning mechanism to find an optimal set of nodes that is topology-independent. Optimal percolation theory has also been applied to identify this optimal set by minimizing the energy of the system [59]. They discovered that the set of optimal influencers, which includes weakly connected nodes often overlooked by previous methods, is much smaller than traditionally identified. Fractal dimension, used to reveal the fractal property of networks, has been applied to identify the key nodes based on their local fractal property at the micro-scale [25]. Its variations have been reviewed [24], shedding light on its applications in influential node identification.

However, these methods typically consider limited topological information, prompting the application of multiple criteria decision-making (MCDM) models to incorporate multiple sources of information in the network. For example, the AHP has been applied to combine information from DC, BC, and CC [61], thereby determining the global score of each node and influential ability. Similarly, evidence theory centrality has been developed to combine information from DC, BC, and CC based on the Dempster–Shafer theory [41], where the weight of each information source is determined by Shannon entropy. Nonetheless, MCDM-based models still face challenges in determining the weights of information sources and constructing multi-level criterion hierarchies to more accurately consider multi-source information.

2.3. Information diffusion models

The influential ability of users is closely related to their ability to propagate information effectively. Hence, the information diffusion model has been widely applied as the benchmark method to evaluate the influential ability of users [21,57,62]. A large number of dynamical models have been proposed to describe the information diffusion process in static networks [63]. The classical one is the threshold model, which was initially developed to characterize collective behaviour and binary decision phenomena and then applied to explore information cascade [64]. Each agent in the network has a threshold and can be one of two states: active or inactive. Initially, only a small number of agents, called seeds, are activated. During the diffusion process, an inactive agent can become active if the fraction of its active neighbours is equal to or larger than its threshold. However, the transition can only occur from the inactive state to the active state, not vice versa. The threshold values [65] and the size of initiators [66] have been extensively studied to understand their impact on the dynamics of information diffusion.

Another one is the independent cascade model that is inspired by interacting particle systems [67,68]. Similar to the threshold model, agents in this model can be in either an active or inactive state. The model assumes that (1) the probability of a node being activated by its active neighbours is independent of the influence of other active nodes, and (2) any active node has only one chance to activate any of its inactive neighbours, regardless of success. The diffusion process follows a similar pattern to the threshold model, but with probabilities assigned to each connection. Generalized cascade models have been proposed to provide more realistic descriptions of the diffusion process [69].

In addition to these models, epidemic models are also commonly used to simulate information propagation in social networks because the information propagation is similar to the epidemic spreading [70]. The original model includes susceptible and infected states, similar to the previous models. Improved versions, such as the Susceptible–Infected–Recovered (SIR) model and Susceptible–Infected–Susceptible model [71], have been developed to capture the contagion process in different scenarios.

However, these dynamical models require a large number of independent repetitions to identify the influential ability of nodes due to the stochastic nature of infection/activation during the information propagation process [63]. Therefore, the dynamical models are not suitable for wide-ranging scenarios, especially for large-scale networks. Nevertheless, they are commonly used as benchmark methods in various disciplines to evaluate the performance of influential ability identification models, particularly the epidemic model. For example, the Susceptible–Infected model, a simple version of the SIR model, was applied to evaluate the performance of the local fuzzy information centrality [62], the random walk-based gravity model [21], and the evidence theory centrality [41]. The SIR model, under different infection rates, has also been employed as the benchmark method to evaluate the performance of several methods in the review paper [13] and advanced models, like the percolation-based evolution optimization framework [57] and gravity-based centrality in hypergraphs [72]. Therefore, the SIR model serves as a benchmark method in this work to compare the performance of different methods based on the studies mentioned.

3. Methodology

In the social network, the influential ability of individuals varies according to the topological structure of the network. Different centrality measures have been developed to consider different structural information, resulting in different evaluations and rankings of the influential ability of nodes. However, how can the most influential ones be identified reasonably, if only a limited number of nodes can be nominated? To address this issue, an evidential reasoning-based influential user evaluation (ERIUE) model is developed in this work to consider multiple sources of structural information. By considering the information from neighbours and paths, coupled with the weights obtained by the feature of each centrality, our developed ERIUE model could identify and rank the influential users in a multifaceted and consistent way.

The influential ability of nodes can be determined by the steps below.

(1) *Criterion hierarchy formulation*: A two-level criterion hierarchy is constructed to consider the multiple sources of structural information from the neighbourhood-based centrality, path-based centrality, and iterative refinement centrality.
(2) *Score and belief distribution evaluation*: The score $\alpha_i(a_l)$ of each node is obtained by different centrality measures considering different structural information, and mapped into a belief distribution $S(\alpha_i(a_l))$ to describe its influential ability after normalization.
(3) *Weight determination*: The weight of each centrality changes according to their own evaluation of the node's influential ability, and it can be obtained by the conflict of belief distribution, the similarity of probability set, and the overlap of evaluations.
(4) *Influential ability identification*: With the determined criterion hierarchy, the influential ability of nodes can then be obtained by our developed model based on the belief distribution of nodes and weight of centrality measures.

The rest of this section is divided into four parts to introduce the ERIUE model in detail.

3.1. Part 1: Criterion hierarchy formulation

In order to consider a variety of structural information to determine the influential nodes in the network $\mathcal{G}(\mathcal{N}, \mathcal{E})$, it is necessary to determine the source of information and the criterion hierarchy to facilitate the fusion of multi-source information. In this work, a two-level criterion hierarchy is constructed in Fig. 1. In the first level, there are three types of centrality measures (criteria), and they are neighbourhood-based centrality C_1, path-based centrality C_2, and iterative refinement centrality C_3. Each type of measure contains three typical methods (sub-criteria) in the second level. Specifically, they are (1) degree $C_{1,1}$, H-index $C_{1,2}$, and k-shell $C_{1,3}$, (2) betweenness $C_{2,1}$, closeness $C_{2,2}$, and gravity centrality $C_{2,3}$, and (3) eigenvector $C_{3,1}$, PageRank $C_{3,2}$, and HITs $C_{3,3}$, which have been introduced in Section 2. The influence ability is mainly evaluated through the network structure in this work, without considering the individual's economic or social attributes, which can be further applied to different networks. The criterion and sub-criterion are both represented as C_i to simplify the process to consider multiple sources of information below, but the criterion hierarchy will not be changed. In this model, $A = \{a_1, a_2, \ldots, a_l, \ldots, a_S\}$ donates the alternative vector, which is the set of nodes $\mathcal{N}$, $E = \{e_1, e_2, \ldots, e_i, \ldots, e_L\}$ is the basic attribute vector that contains all centrality measures, and $H = \{H_1, H_2, \ldots, H_n, \ldots, H_N\}$ is the set of evaluation grades.

3.2. Part 2: Score and belief distribution evaluation

For each centrality $i \in [1, 9]$ (as there are nine methods), the score of node $l \in \mathcal{N}$ can be represented by $\alpha_i(a_l)$, and the way of obtaining it has been introduced in Section 2.1. However, these values fall in different ranges due to differences in method design. Therefore, it needs to be normalized before further processing. The normalized value $\hat{\alpha}_i(a_l)$ can be obtained by,

$$\hat{\alpha}_i(a_l) = \frac{\alpha_i(a_l) - \min_{a_l} \alpha_i(a_l)}{\max_{a_l} \alpha_i(a_l) - \min_{a_l} \alpha_i(a_l)}, \tag{12}$$

to avoid the effect of the different ranges of scores. In this case, the influential ability of node l evaluated by centrality i is quantitatively

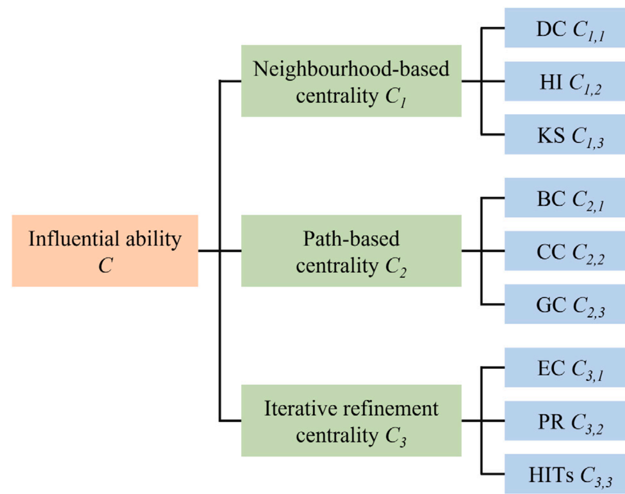


Fig. 1. The two-level criterion hierarchy for identifying influential nodes.

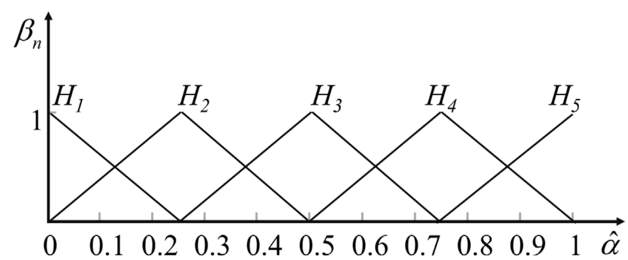


Fig. 2. Triangular fuzzy membership function that maps the normalized score to the degree of belief in the belief distribution.

Table 1
Relationship between evaluation grade H_n and corresponding normalized score $\hat{\alpha}$.

Evaluation grade	Linguistic term	Normalized score $\hat{\alpha}$
H_1	Very uninfluential	$[0, 0.25]$
H_2	Uninfluential	$[0, 0.5]$
H_3	Fair	$[0.25, 0.75]$
H_4	Influential	$[0.5, 1]$
H_5	Very influential	$[0.75, 1]$

described by the normalized score $\hat{\alpha}_i(a_l)$. For the same score, it is sometimes rated as *influential* and sometimes as *very influential*, thus, the score could be converted into the degree of belief of belonging to a certain linguistic item describing its influential ability. In this work, the degree of belief $\beta_{n,i}(a_l)$ is obtained by the triangular fuzzy membership function, where the map function is shown in Fig. 2. The relationship between evaluation grade and the corresponding normalized score is shown in Table 1. With the degree of belief, the belief distribution is shown as,

$$\begin{aligned} S\left(e_i\left(a_l\right)\right) = \{ & \left(H_n, \beta_{n,i}\left(a_l\right)\right), n = 1, 2, \ldots, N; \\ & \left(H_H, \beta_{H,i}\left(a_l\right)\right) \}, \end{aligned} \tag{13}$$

where $N = 5$ in this work, H_n is the nth evaluation grade and $\beta_{n,i}(a_l)$ indicates the degree of belief that node l is evaluated as H_n by centrality i. H_{n+1} is usually assumed to be preferred to H_n [73]. Notably, $0 \le \beta_{n,i}(a_l) \le 1$ and $\sum_{n=1}^{N} \beta_{n,i}(a_l) \le 1$. The assessment $S\left(e_i\left(a_l\right)\right)$ is uncompleted when $\sum_{n=1}^{N} \beta_{n,i}(a_l) < 1$, and completed when it equals to 1. This work assumes that the initial judgement is completed, that is, $\sum_{n=1}^{N} \beta_{n,i}(a_l) = 1$, leading to the remaining mass unassigned to any grade $\beta_{H,i}(a_l) \equiv 0$ for any centrality i. The subjectiveness of each centrality can be described by the framework of belief distributions, enabling all information to be reasonably considered.

3.3. Part 3: Weight determination

A single information source may have biased influential ability evaluation results, thus, multiple sources of information need to be considered in this work to fill this important gap. For the multi-criteria decision-making process, the most important step is to assign the weight of each information source (centrality measure here) to evaluate its rationality because each centrality may give different scores and rankings to nodes. Usually, the more reliable the information source is, the higher its weight value. Three types of information related to the score and belief distribution obtained by each centrality are applied to determine its weight.

Since the structure of nodes in the network is usually different, regardless of the local or the global structure, a centrality should give fewer nodes the same score, thereby distinguishing their influential ability. Therefore, the overlap of scores should be considered to determine the weight of the centrality. The score of node l evaluated by centrality i is represented by $\hat{\alpha}_i(a_l)$, and the scores of all nodes can be represented by a probability set P_i, which is obtained by

$$\begin{aligned} P_i &= \{p_i(a_1), \dots, p_i(a_l), \dots, p_i(a_S)\} \\ &= \{\hat{\alpha}_i(a_1)/\sum_l \hat{\alpha}_i(a_l), \dots, \hat{\alpha}_i(a_S)/\sum_l \hat{\alpha}_i(a_1)\}. \end{aligned} \tag{14}$$

The overlap of scores in centrality i can be then determined based on the Shannon entropy,

$$E_i = -\sum_{l=1}^{S} p_i\left(a_l\right) \log_2 p_i\left(a_l\right), \tag{15}$$

and the overlap-based weight w_e^i of centrality i can be obtained by,

$$w_e^i = \frac{E_i}{\sum_i E_i}, \tag{16}$$

which satisfies $0 \leq w_e^i \leq 1$ and $\sum_i w_e^i = 1$.

In addition to considering the score obtained by the centrality itself, it is also necessary to consider the scores obtained by all centralities to determine whether the evaluation result of this centrality is reasonable. This is very common in group decision-making. For example, when an expert gives a different (or diametrically opposite) opinion from all other experts, its weight in the decision-making process will be relatively lower than others. Similarly, if a node evaluated as low (or high) influential by other centralities is identified as a high (or low) influential one by this centrality, its weight will be lower than other centralities. The Jensen–Shannon divergence (relative entropy) is a commonly used method for evaluating the difference between a pair of probability sets in information theory, thus, the difference between the results obtained by centrality i and j can be obtained by,

$$\gamma_{ij} = \frac{1}{2}\left(\sum_l p_i\left(a_l\right)\log_2\frac{p_i\left(a_l\right)}{p_c\left(a_l\right)} + \sum_l p_j\left(a_l\right)\log_2\frac{p_j\left(a_l\right)}{p_c\left(a_l\right)}\right), \tag{17}$$

where $p_c(a_l)$ is the element in the probability set $P_c = \left(P_i + P_j\right)/2$. The length of all probability sets is the same, that is, the number of nodes $|\mathcal{N}|$ in the network. The divergence $\gamma_{ij} \in [0,1]$ between a pair of centralities is divergence. The value of γ_{ij} is larger when the results obtained by the two centralities differ more. However, a centrality is given less weight when it obtains a result that is less similar to all other centralities. Hence, the similarity-based weight w_s^i is obtained by,

$$w_s^i = \frac{\sum_{j\neq i} 1-\gamma_{ij}}{\sum_i \sum_{j\neq i} 1-\gamma_{ij}}. \tag{18}$$

Similarly, $0 \leq w_s^i \leq 1$ and $\sum_i w_s^i = 1$.

The score of nodes can indicate the difference between centrality measures, and the belief distribution expressing the influential ability of nodes can express the conflict between two centralities from the alternative level. Based on the subjective judgments, a centrality will be assigned a higher weight when it has less conflict with other measures (attributes) in describing the influence (in the evaluation grade level) of each node (in the alternative level). The conflict [74] between centrality i and j on node l can be measured by,

$$\begin{aligned} D\left(e_{ij}\left(a_l\right)\right) = &\frac{1}{u\left(H_N\right)-u\left(H_1\right)}\times \\ &\sum_{n=1}^{N-1}\sum_{s=n+1}^{N}\beta_{n,|i-j|}\left(a_l\right)\beta_{s,|i-j|}\left(a_l\right)u\left(H_{|s-n|}\right), \end{aligned} \tag{19}$$

where $u\left(H_n\right)$ indicates the utility of H_n that follows $u\left(H_{n+1}\right) > u\left(H_n\right)$, and $1/\left(u\left(H_N\right)-u\left(H_1\right)\right)$ is the normalization factor such that the maximum value of $D\left(e_{ij}\left(a_l\right)\right)$ is 1. In addition, $\beta_{n,|i-j|}\left(a_l\right) = \left|\beta_{n,i}\left(a_l\right)-\beta_{n,j}\left(a_l\right)\right|$ indicates the belief degree of dissimilarity and $u\left(H_{|s-n|}\right) = u\left(H_s\right)-u\left(H_n\right)$ describes the difference in the utility of the different evaluation grades. The conflict between the centrality i and the other centralities on all nodes can be obtained,

$$\varphi_i = \frac{\sum_{j\neq i}\sum_l D\left(e_{ij}\left(a_l\right)\right)}{(L-1)\times S}, \tag{20}$$

where L and S are the lengths of the attribute vector and alternative vector, respectively. The conflict-based weight w_c^i of centrality i can be obtained by,

$$w_c^i = \frac{1-\varphi_i}{L-\sum_i \varphi_i}, \tag{21}$$

because the weight is larger when the conflict is smaller. It also follows $0 \leq w_c^i \leq 1$ and $\sum_i w_c^i = 1$.

By combining the information from the three perspectives, the overall weight of each centrality will be obtained by,

$$w^i = \frac{w_e^i + w_s^i + w_c^i}{3}, \tag{22}$$

which is the average value of the overlap-based, similarity-based, and conflict-based weights.

3.4. Part 4: Influential ability identification

Based on the obtained belief distributions (Eq. (13)) and weights (Eq. (22)), the ER approach is applied to aggregate the multiple sources of information with weights and obtain the overall belief distribution describing the influential ability of nodes. The basic probability mass $m_{n,i}$ describing the belief assigned to H_n by the centrality i can be obtained by,

$$\begin{aligned} m_{n,i} &= w^i \times \beta_{n,i}, \\ m_{H,i} &= 1 - w^i \times \sum_n \beta_{n,i}, \end{aligned} \tag{23}$$

where $m_{H,i}$ indicates the remaining mass that cannot be assigned to any evaluation grade H_n by the centrality i, which can be decomposed into two parts:

- $\bar{m}_{H,i} = 1 - w^i$ is bounded by the relative importance of centrality i, indicating the degree to which other centralities can contribute to the assessment;
- $\tilde{m}_{H,i} = w^i\left(1-\sum_n \beta_{n,i}\right)$ is caused by the incompleteness of information from centrality i.

According to the criterion hierarchy in Fig. 1, the basic probability mass on sub-criteria $C_{i,j}$ is aggregated to the criterion C_i at the first level based on the recursive ER algorithm [73], and then aggregated to the top-level decision criterion C again with the same process (including weight determination and aggregation). Let the probability mass $m_{n,I(i)}$ be the degree to which the first $I(i)$ attributes support evaluation grade H_n and $m_{H,I(i)}$ be the remaining mass that is unassigned to any evaluation grade by the first $I(i)$ attributes. The initial setting is

$m_{n,I(1)} = m_{n,1}$ and $m_{H,I(1)} = m_{H,1}$. The combined probability mass can be obtained by,

$$
\begin{aligned}
m_{n,I(i+1)} &= K_{I(i+1)} \left(m_{n,I(i)} m_{n,i+1} + m_{n,I(i)} m_{H,i+1} + m_{H,I(i)} m_{n,i+1} \right), \\
\tilde{m}_{H,I(i+1)} &= K_{I(i+1)} \left(\tilde{m}_{H,I(i)} \tilde{m}_{H,i+1} + \bar{m}_{H,I(i)} \tilde{m}_{H,i+1} + \tilde{m}_{H,I(i)} \bar{m}_{H,i+1} \right), \\
\bar{m}_{H,I(i+1)} &= K_{I(i+1)} \left(\bar{m}_{H,I(i)} \bar{m}_{H,i+1} \right), \\
K_{I(i+1)} &= \left(1 - \sum_{t} \sum_{j \neq t} m_{t,I(i)} m_{j,i+1} \right)^{-1},
\end{aligned} \tag{24}
$$

for $i = 1, 2, \ldots, L$. After all information is aggregated, the combined degree of belief is obtained by,

$$
\begin{aligned}
\beta_n &= \frac{m_{n,I(L)}}{1 - \bar{m}_{H,I(L)}}, \\
\beta_H &= \frac{\tilde{m}_{H,I(L)}}{1 - \bar{m}_{H,I(L)}},
\end{aligned} \tag{25}
$$

where β_n is the combined degree of belief that a node is evaluated as H_n, and β_H is the combined degree of belief that is not assigned to any grade, indicating the degree of incompleteness in the overall assessment. The overall assessment of the influential ability of node l is represented by the belief distribution,

$$
\begin{aligned}
S\left(C\left(a_l\right)\right) = \{ & \left(H_n, \beta_n\left(a_l\right)\right), n = 1, 2, \ldots, N; \\
& \left(H_H, \beta_H\left(a_l\right)\right) \},
\end{aligned} \tag{26}
$$

The influential ability of node l is then quantified by,

$$
\zeta_l = \sum_{n=1}^{N} \omega_n \beta_n(a_l), \tag{27}
$$

where $\omega_n \in \omega_{N \times 1}$ is the weight parameter describing the contribution of each evaluation grade to the influential ability, which is defined as the utility of each grade in the ER approach. The pseudo-code of our developed ERIUE model based on the formulated criterion hierarchy is shown in Algorithm 1. After considering the evidence from each centrality measure, the time complexity of the ER-based centrality model is primarily determined by two factors: the number of criteria in the decision-making process (L) and the number of nodes in the network (S). The decision-making process involves aggregating the evidence from each centrality measure and determining the weight of each measure based on the specified criteria. Additionally, the model needs to iterate over all nodes in the network to combine the evidence to determine its influential ability. Therefore, the time complexity of the ERIUE model can be approximated as $O(L \times S)$. This complexity provides a measure of the algorithm's efficiency and scalability, enabling its application to larger networks with a reasonable computational cost.

Algorithm 1: Evidential reasoning-based influential users evaluation model.

Input: Network structure $\mathcal{G}(\mathcal{N}, \mathcal{E})$; Mapping rule $f(x)$.
Output: Influential ability ζ_l of node l.

1 Formulate the criterion hierarchy in Fig. 1.
`/* Aggregate information from sub-criteria` $C_{i,j}$ `to the criterion` C_i`. */`
2 **for** *Criterion* C_i **do**
3   **for** *Sub-criterion* $C_{i,j}$ **do**
    `/* Evaluate the score. */`
4   **for** *Node* l **do**
5   Obtain the score $\alpha_{i,j}(a_l)$ for node l by the selected centrality (sub-criterion $C_{i,j}$) from the network structure.
6   Normalize the scores $\hat{\alpha}_{i,j}(a_l)$ by Eq. (12).
    `/* Evaluate the belief distribution. */`
7   **for** *Node* l **do**
8   **for** *Grade level* n **do**
9   Map the score to the belief degree by $\beta_{n,i,j} = f\left(\hat{\alpha}_{i,j}(a_l)\right)$ shown in Fig. 2.
10   Get the degree distribution $S(e_{i,j}(a_l)) = \left\{\left(H_n, \beta_{n,i,j}(a_l)\right), n = 1, 2, ..., N, \left(H_H, \beta_{H,i,j}(a_l)\right)\right\}$ in Eq. (13).
  `/* Determine the weight for each centrality. */`
11   **for** *Sub-criterion* $C_{i,j}$ **do**
12   Obtain the overlap-based weight of each centrality $w_e^{i,j}$ by the Shannon entropy shown in Eq. (16).
13   **for** *Other sub-criteria* $C_{i,k}$ **do**
14   Measure the differences in the probability sets γ_{jk} and obtain the similarity-based weight $w_s^{i,j}$ by Eq. (18).
15   Evaluate the conflict between centralities $D\left(e_{jk}\left(a_l\right)\right)$ and obtain the conflict-based weight $w_c^{i,j}$ by Eq. (21).
16   Obtain the weight for each centrality $w^{i,j}$ by Eq. (22).
  `/* Get the aggregated information. */`
17   **for** *Node* l **do**
18   Aggregate information from sub-criteria $C_{i,j}$ to criterion C_i by the recursive ER approach.
19   Obtain the aggregated belief distribution $S\left(C_i\left(a_l\right)\right)$ for node l shown in Eq. (26).
`/* Aggregate information from criteria` C_i `to the top-level decision criterion` C `and identify the influential ability of each node. */`
20 **for** *Criterion* C_i **do**
21   Repeat Lines 11 – 16 to determine the weight w^i of each criterion from the three perspectives.
22 **for** *Node* l **do**
23   Aggregate information from C_i to C by the recursive ER approach.
24   Obtain the overall belief distribution $S\left(C\left(a_l\right)\right) = \{\left(H_n, \beta_n\left(a_l\right)\right), n = 1, 2, \ldots, N; \left(H_H, \beta_H\left(a_l\right)\right)\}$ for node l.
25   Identify the influential ability of users ζ_l by Eq. (27).

4. Evaluations

4.1. Datasets

In order to evaluate the performance and applicability of our proposed model, three commonly used real-world social networks are used as examples. The first one is the jazz network [75] which describes the collaboration between jazz musicians. The second one is the coauthor network [76] which describes the scientific collaboration between network scientists. In this network, there is an edge between two scientists when they published papers together. The last one is the blog network [77] that describes the hyperlinks between political blogs in the USA. The topological properties of the three networks are shown in Table 2. $|\mathcal{N}|$ and $|\mathcal{E}|$ indicate the number of nodes and edges, respectively, $\langle k \rangle$ and $\langle d \rangle$ represent the average degree and average distance, respectively, C and r are the clustering coefficient and the assortative coefficient.

4.2. Top-10 nodes

To illustrate the applicability, the IDs of top-10 nodes in the three real-world networks identified by our proposed model and comparison measures are shown in Table 3. In this work, the weight parameters are $\omega = \{0.1, 0.25, 0.5, 0.75, 1\}$, which is used as an example to show the

Table 2
Topological properties of networks [51].

Network	$\lvert\mathcal{N}\rvert$	$\lvert\mathcal{E}\rvert$	$\langle k\rangle$	$\langle d\rangle$	C	r
Jazz	198	2472	27.6970	2.2350	0.6334	0.0202
Coauthor	379	914	4.8232	6.0419	0.7981	−0.0817
Blog	1222	16714	27.3552	2.7375	0.3600	−0.2213

performance of the ERIUE algorithm. For the jazz network, node 136 is identified as the most influential one because six of the nine centrality measures support this point. There are only three measures to list node 60 as the most influential node, so it is ranked second by the proposed model. As for nodes 132 and 168, six measures hold the opinion that node 132 is more influential and three measures even rank it as the second place, so they rank third and fourth in our proposed model. Furthermore, the top four nodes obtained by our proposed model are exactly the same as that listed by another four centrality measures. As for the rest of the nodes, they are all obtained from the perspective of multi criteria decision-making, that is, one node will be ranked higher by our proposed model only when it is evaluated as having higher influential ability by more measures. For the coauthor network, nodes 4 and 5 are identified as the two most influential nodes by our proposed method as most centrality measures support this. As for node 26, it is even slightly worse than node 16, as it is not listed in the top 10 nodes by four measures (even it is listed as the most influential node by three measures), while node 16 is not listed in the top-10 nodes by only two measures. The situation is similar in the blog network. Node 12 is listed as the most influential by four measures and the second most influential by another four measures, thus, it is identified as the most influential by the ER centrality. Node 28 is another node rated well by most measures. More details about the top-10 nodes can be found in Table 3. By considering various criteria, our proposed ERIUE algorithm aims to identify influential nodes based on their collective recognition across different algorithms. It leverages the strengths of different centrality measures and combines them in a way that captures the overall influential ability of nodes more accurately. In contrast, methods that solely rely on a single type of structural information or centrality measure may overlook certain aspects of node influence. Overall, the ERIUE algorithm's consideration of multi-source structural information contributes to its perceived reasonability and effectiveness in identifying influential nodes in social networks.

4.3. Susceptible–infected–recovered model

The list of top-10 nodes obtained by different centrality measures simply illustrates the applicability of our proposed algorithm. The performance of each centrality is then evaluated by comparing them with the benchmark method. As discussed in Section 2.3, information diffusion models are commonly used as benchmark methods for assessing the influential ability of nodes in social networks, with the SIR model being particularly prominent. Following the recommendations put forth by previous studies [13,21,57], the classical SIR model is used as the benchmark method in this work to provide further insights into the performance of each centrality measure.

To evaluate the influential ability of node i, only node i is selected as the initially infected node, while all other nodes are considered susceptible. At each time t, infected nodes have the potential to either infect susceptible neighbour nodes with probability β or recover from the infection without engaging in subsequent propagation with recovery probability λ. This process is captured by the following set of ordinary differential equations,

$$\begin{cases} \frac{ds(t)}{dt} = -\beta s(t)i(t) \\ \frac{di(t)}{dt} = \beta s(t)i(t) - \lambda i(t) \\ \frac{dr(t)}{dt} = \lambda i(t) \end{cases}, \tag{28}$$

where $s(t)+i(t)+r(t) \equiv |\mathcal{N}|$ holds for all time t. This simulation proceeds until reaching a predetermined time point t'. The influential ability of node i is quantified by the ratio of the number of recovered nodes N_r to the total number of nodes in the network $|\mathcal{N}|$. Mathematically, we express this as $F_i = N_r/|\mathcal{N}|$. Given that the infection process is inherently stochastic and dependent on the probabilities β and λ, the benchmark influential ability of node i is the average F_i of 500 independent repeated numerical experiments. The benchmark ranking list σ is then determined by the corresponding values of F_i for each node. In this work, the probabilities are set according to classical experiments: β reaches the epidemic threshold $\beta_c \approx \frac{\langle k\rangle}{\langle k^2\rangle - \langle k\rangle}$ [78], $\lambda = 1$, and $t' = 40$. In order to evaluate the correlation between the list obtained by centrality measures and that obtained by the benchmark method, Kendall's Tau correlation coefficient is applied in this work,

$$\tau = \frac{2(n_+ - n_-)}{|\mathcal{N}|(|\mathcal{N}| - 1)}, \tag{29}$$

where n_+ and n_- represent the number of concordant and discordant pairs, respectively. A centrality measure will get a higher value of τ if it can obtain a ranking list that is closer to the list obtained by the benchmark method. The coefficient τ in the three networks is shown in Table 4, where the highest correlation coefficients are highlighted in bold.

In the analysis of the three networks, it can be found that our proposed ERIUE model consistently exhibits the highest correlation with the benchmark method. This implies that irrespective of the network type, the ERIUE model consistently yields a ranking list that closely resembles the benchmark, owing to its comprehensive consideration of network topology information. On the other hand, the performance of the comparative centralities is unstable in different networks. For example, τ obtained by HITs closely aligns with our developed method in the jazz and blog networks, but ranks significantly lower in the coauthor network. This demonstrates that our proposed model has robust performance across diverse application scenarios. With its ability to incorporate multi-source network topology information, the ERIUE algorithm consistently achieves evaluation results that closely align with the benchmark, highlighting the superiority of our proposed algorithm.

To account for the variability of the infection probability β in real-world scenarios, the performance of each method is further explored within the range of $\beta \in [0.5\beta_c, 1.5\beta_c]$. It should be noted that the benchmark ranking list may differ under different β values. The Kendell correlation coefficient τ is still applied to measure the correlation between the list obtained by the centralities and the benchmark with varying β. A higher τ value indicates that the ranking list generated by a method closely aligns with the benchmark ranking list under a specific β, signifying the effectiveness of the method in this context.

The results are shown in Fig. 3. In the jazz network, τ obtained by the ER centrality is consistently the highest across different values of β/β_c [Fig. 3 (a)]. Some measures exhibit slightly higher or close τ values to our developed model at specific β values, such as DC and GC at $\beta = 0.5\beta_c$ and EC and HITs at $\beta = 1.2\beta_c$. However, for the majority of cases, our proposed model consistently outperforms the comparative centrality measures in terms of τ values. This indicates that our proposed method consistently achieves results that closely resemble the benchmark ranking list under varying infection probabilities. The case is similar in the coauthor and blog networks. In the coauthor network, only τ obtained by GC at $\beta = 0.5\beta_c$ is slightly higher than that obtained by our proposed model [Fig. 3 (b)]. In the blog network, some measures can achieve τ close to our developed model at $\beta = 0.5\beta_c$ [Fig. 3 (c)]. However, overall, our proposed method consistently exhibits higher correlation coefficients across varying β values.

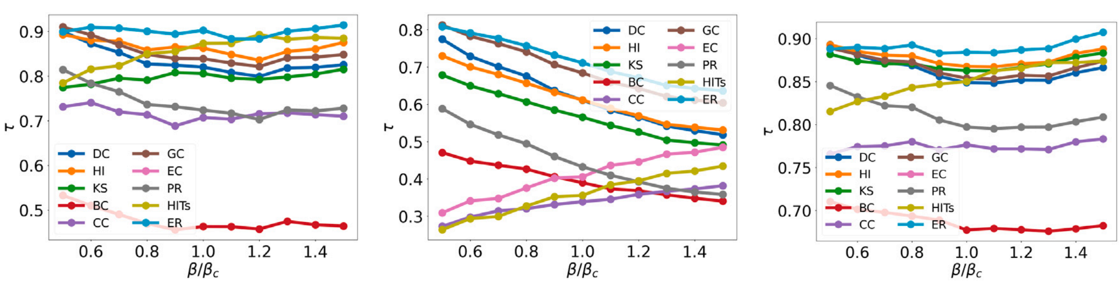
Through the adequate consideration and rational fusion of information, the ERIUE algorithm consistently yields results that closely align with the benchmark method across various networks and parameter settings, demonstrating the applicability of our approach to diverse scenarios. This sets our method apart from other approaches

Table 3
The list of top-10 nodes identified by different centrality measures.

Network	Rank	DC	HI	KS	BC	CC	GC	EC	PR	HITs	Proposed (ER)
Jazz	1	136	136	60	136	136	136	60	136	60	136
	2	60	60	132	153	60	60	132	60	132	60
	3	132	132	168	60	168	132	136	168	136	132
	4	168	168	99	149	70	168	168	132	168	168
	5	70	99	108	168	83	70	108	149	108	108
	6	99	108	131	167	132	108	99	70	99	99
	7	108	131	122	189	122	99	131	83	131	70
	8	83	194	100	115	194	83	70	167	70	131
	9	158	70	101	96	174	158	83	96	83	83
	10	7	83	98	83	158	194	194	158	194	122
Coauthor	1	4	4	4	26	26	4	4	26	4	4
	2	5	5	5	51	95	5	5	4	5	5
	3	26	16	16	169	51	26	16	5	16	16
	4	16	15	15	95	231	16	15	95	15	26
	5	67	51	45	67	100	95	45	67	45	15
	6	70	45	46	5	52	67	46	16	46	95
	7	95	46	47	231	5	51	47	51	47	45
	8	15	47	176	100	44	15	176	32	176	51
	9	51	176	177	44	234	231	177	70	177	46
	10	113	177	70	66	297	70	250	8	250	47
Blog	1	12	28	67	304	28	12	12	304	12	12
	2	28	67	12	12	12	28	14	12	14	28
	3	304	12	14	94	16	14	16	94	16	16
	4	14	14	16	28	14	16	67	28	67	14
	5	16	16	18	16	36	304	52	16	52	304
	6	94	18	52	145	67	67	18	14	18	67
	7	6	52	176	14	94	94	28	6	28	52
	8	67	176	47	300	35	6	47	35	47	94
	9	35	47	4	35	145	35	73	145	73	6
	10	145	4	161	67	304	36	9	67	9	18

Table 4
Kendall's Tau correlation coefficient between the lists obtained by centrality measures and that obtained by the benchmark method.

	DC	HI	CC	BC	KC	GC	EC	PR	HITs	ER
Jazz	0.8202	0.8626	0.8058	0.4632	0.7074	0.8389	0.8732	0.7238	0.8732	**0.9026**
Coauthor	0.6110	0.6115	0.5654	0.3898	0.3389	0.6841	0.4057	0.4326	0.3557	**0.7113**
Blog	0.8489	0.8678	0.8629	0.6774	0.7765	0.8540	0.8503	0.7971	0.8502	**0.8843**



(a) Jazz (b) Coauthor (c) Blog

Fig. 3. Kendall's Tau correlation coefficient τ under different values of β.

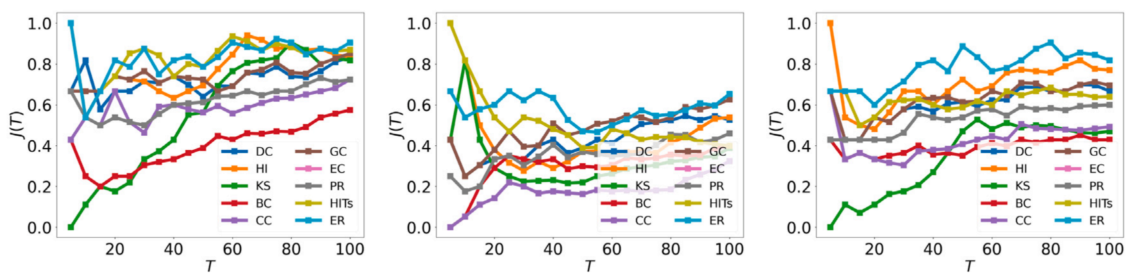
that only consider a single type of structural information. This observation underscores the significance of incorporating comprehensive structural information for analysing the influential ability of users in social networks.

In addition to the Kendall correlation coefficient, which measures the correlation between full ranking lists, it is also important to consider the similarity between the top-ranked nodes, as these are often of greater interest due to their high influential ability. To assess this, the Jaccard similarity coefficient [79] is applied to measure the similarity between the top-T nodes in the ranking lists obtained by the benchmark method and the centrality measures,

$$J(T) = \frac{|\sigma(T) \cap R(T)|}{|\sigma(T) \cup R(T)|}, \tag{30}$$

where $\sigma(T)$ and $R(T)$ are the top-T nodes in the benchmark ranking list and the ranking list obtained by different centralities, respectively. Here, T grows from 5 to 100 with an intervals of 5. The values of $J(T)$ in the three networks with different values of T are shown in Fig. 4. Similar to τ, a higher value of $J(T)$ indicates a more effective method.

In the jazz network, the ER centrality consistently achieves the highest $J(T)$ values in most cases, although there are instances where other measures yield higher values [Fig. 4 (a)]. In the coauthor network, several centrality measures, such as HI, CC, EC, and HITs, outperform the ER centrality when $T \leq 15$ [Fig. 4 (b)]. Similarly, in the blog network, HI achieves higher values of $J(T)$ than the ER centrality for $T \leq 15$ [Fig. 4 (c)]. However, as T increases beyond 15, the ER centrality consistently outperforms other measures, indicating its strong ability to identify influential users. Furthermore, it is worth noting that different centrality measures have significantly different performances



(a) Jazz (b) Coauthor (c) Blog

Fig. 4. Jaccard similarity coefficient $J(T)$ under different values of T.

in different networks, underscoring the limitations of these comparison methods. Conversely, the ERIUE algorithm consistently approaches the benchmark results across various scenarios, demonstrating its stability and real-world applicability.

In this work, the Kendell correlation coefficient and Jaccard similarity coefficient are utilized to evaluate the performance of different centrality measures in assessing the influential ability of nodes. These metrics provide insights into the overall correlation of ranking lists and the similarity of top-ranked nodes, respectively, shedding light on the effectiveness of each centrality measure. By analysing the results across different values of β and T in the three networks, we have observed that our proposed ERIUE algorithm consistently achieves the closest results to the benchmark, outperforming the comparative centrality measures. The stability and superiority of the ERIUE algorithm are crucial for its broader application, which cannot be matched by the comparative centrality measures that consider only a single type of structural information. This demonstrates the practical relevance and applicability of our algorithm in real-world scenarios.

4.4. Impact of weight

In this study, the impact of weight determination on the identification of influential nodes is investigated by conducting an ablation study. Three cases are considered, each involving a combination of two types of factors to determine the weight of each criterion.

- **Case 1**: Only consider w_c^i and w_s^i.
- **Case 2**: Only consider w_c^i and w_e^i.
- **Case 3**: Only consider w_s^i and w_e^i.

In all cases, the weight of each criterion is the average of the two considerations. To examine the impact of weight determination, two networks are selected as examples, and the ranking lists are obtained in the three cases, along with the original case that considers all three factors simultaneously. The performance of each case is evaluated using the Kendell correlation coefficient τ and the Jaccard similarity coefficient $J(T)$ (Fig. 5).

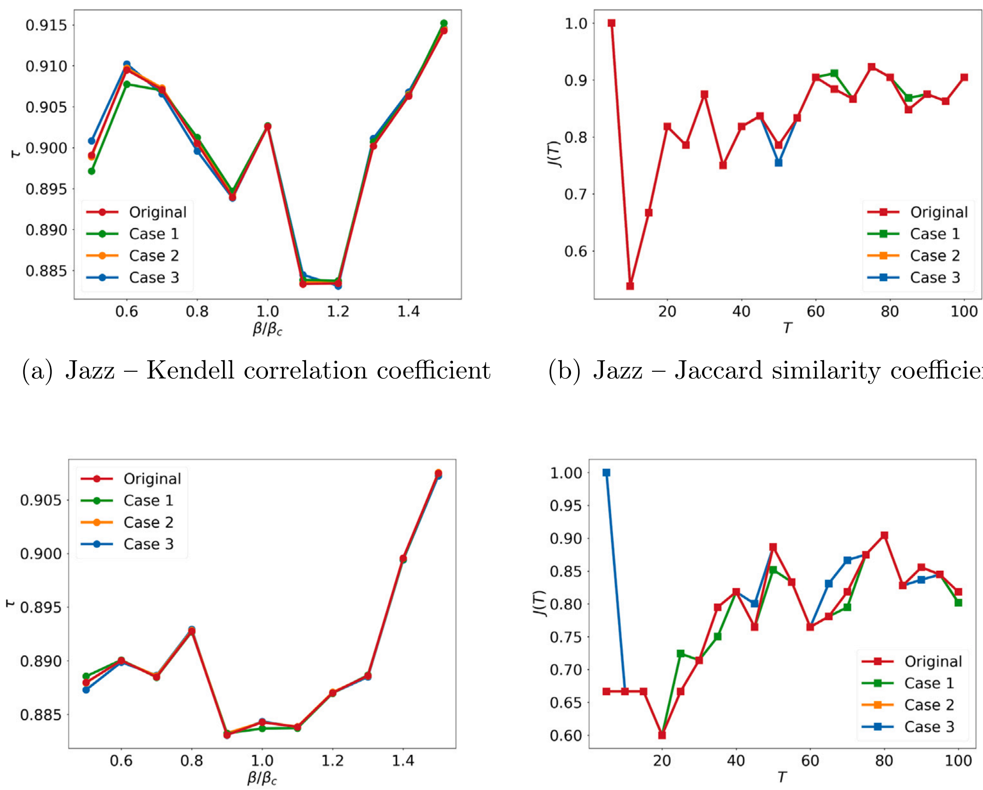
The results show that the difference in τ for the both networks is small, except for cases when $\beta \leq \beta_c$ [Fig. 5 (a) and (c)]. This suggests that the overall ranking of influential nodes remains largely unaffected by how the weights are determined, except under low infection probabilities. When considering the similarity of top-ranked nodes, the impact of weight determination becomes more pronounced. In the jazz network, the values of $J(T)$ are the same for $T < 50$ but become different for $T \geq 50$ [Fig. 5 (b)]. This indicates that nodes with lower influential ability may be ranked differently depending on the weight determination. Similarly, in the blog network, Case 3 achieves a significantly higher $J(5)$ value compared to the other cases, implying that this weight determination approach successfully identifies the same top five highly influential users as the benchmark method. For other values of T, the results are similar, with identical $J(T)$ values for $T \leq 20$ and different values for $T > 20$ [Fig. 5 (d)].

In conclusion, the ablation study demonstrates that the determination of weights has a limited impact on the overall ranking of influential nodes, but may result in variations for nodes with lower influential ability. Nevertheless, this does not compromise the algorithm's ability to identify high-impact users, which is the primary focus of this study. It is important to note that the determination of weights is flexible and can be adjusted according to specific needs in different scenarios, extending the scalability of the algorithm beyond the three factors defined in this study. The flexibility in weight determination allows for customization based on specific scenarios, highlighting the algorithm's versatility and adaptability.

5. Conclusion

In this paper, an original and novel ERIUE model for identifying influential users is proposed by considering multiple sources of structural information with weights. Unlike existing centrality measures that focus on a single type of structural information, the proposed model considers different types of information through the recursive ER approach. In this model, a two-level criterion hierarchy is formulated to consider different information through several existing centrality measures. Each typical centrality evaluates the influential ability based on its feature, and the scores are mapped to a belief distribution after normalization. A comprehensive weight assignment way is then developed in this work to evaluate the results obtained by each centrality. Specifically, a centrality will be assigned higher weight if (1) fewer nodes are given the same score by this centrality; (2) its scores (expressed as the probability set) are more similar to the scores obtained by other measures; and (3) its belief distribution has less conflict with that of other measures. The three factors are averaged to determine the weight of each centrality. The basic probability mass is obtained by the belief distribution and corresponding weight, which are further used to aggregate information in the recursive ER approach. After a two-stage aggregation of information based on the criterion hierarchy, the influential ability of nodes is evaluated based on the overall assessment. This model provides a more comprehensive evaluation of influential ability by considering different types of structural information from the perspective of multiple criteria decision-making.

Three real-world social networks are used to evaluate our proposed ERIUE model. The top 10 node lists obtained by various centrality measures are used to illustrate the principle of this model, which is to adopt the ranking relationship supported by more centralities. The SIR model is then applied as the benchmark to evaluate the performance of each centrality. In the numerical simulation, the correlation is measured by Kendall's Tau correlation coefficient, and the similarity



(a) Jazz – Kendell correlation coefficient (b) Jazz – Jaccard similarity coefficient

(c) Blog – Kendell correlation coefficient (d) Blog – Jaccard similarity coefficient

Fig. 5. The impact of weight in two examples networks, quantified by τ and $J(T)$.

between a portion of top-ranked nodes is measured by the Jaccard similarity coefficient. The experimental results on the three social networks demonstrate the applicability and superiority of our developed model. In the future, this work can serve as a framework to incorporate additional types of information to identify influential nodes in various types of networks, such as hubs in transportation networks and key electrical elements in power grid networks.

CRediT authorship contribution statement

Tao Wen: Conceptualization, Formulation, Methodology, Simulation, Writing – review & editing. **Yu-wang Chen:** Supervision, Formulation, Methodology, Reviewing & editing. **Tahir abbas Syed:** Evaluation, Reviewing & editing. **Ting Wu:** Evaluation, Reviewing & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We thank the three anonymous reviewers and editors for generously aiding us in enhancing the presentation of our work. The work is partially supported by the Economic and Social Research Council (ESRC), UK as part of the Centre for Doctoral Training in Data Analytics and Society.

Appendix

List of Abbreviations

MCDM	Multiple criteria decision-making
AHP	Analytic Hierarchy Process
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
ER	Evidential reasoning
ERIUE	Evidential reasoning-based influential users evaluation
SIR	Susceptible–Infected–Recovered
DC	Degree centrality
KS	*k*-shell
HI	H-index
CC	Closeness centrality
BC	Betweenness centrality
GC	Gravity-based centrality
EC	Eigenvector centrality
PR	PageRank
HITs	Hyperlink-Induced Topic Search

References

[1] Newman M. Networks. Oxford University Press; 2018.
[2] Scott J. Social network analysis. Sociology 1988;22(1):109–27.
[3] Lü L, Zhou T. Link prediction in complex networks: A survey. Physica A 2011;390(6):1150–70.
[4] Keskin BB, Griffin EC, Prell JO, Dilkina B, Ferber A, MacDonald J, et al. Quantitative investigation of wildlife trafficking supply chains: A review. Omega 2022;102780.
[5] Lü L, Medo M, Yeung CH, Zhang Y-C, Zhang Z-K, Zhou T. Recommender systems. Phys Rep 2012;519(1):1–49.
[6] Resnick P, Varian HR. Recommender systems. Commun ACM 1997;40(3):56–8.

[7] Liao H, Li X, Tang M. How to process local and global consensus? A large-scale group decision making model based on social network analysis with probabilistic linguistic information. Inform Sci 2021;579:368–87.
[8] Tang M, Liao H. From conventional group decision making to large-scale group decision making: What are the challenges and how to meet them in big data era? A state-of-the-art survey. Omega 2021;100:102141.
[9] Wen T, Cao J, Cheong KH. Gravity-based community vulnerability evaluation model in social networks: GBCVE. IEEE Trans Cybern 2023;53(4):2467–79.
[10] Zhao J, Cheong KH. Obfuscating community structure in complex network with evolutionary divide-and-conquer strategy. IEEE Trans Evol Comput 2023.
[11] Óskarsdóttir M, Bravo C. Multilayer network analysis for improved credit risk prediction. Omega 2021;105:102520.
[12] Wen T, Gao Q, wang Chen Y, Cheong KH. Exploring the vulnerability of transportation networks by entropy: A case study of Asia–Europe maritime transportation network. Reliab Eng Syst Saf 2022;108578.
[13] Lü L, Chen D, Ren X-L, Zhang Q-M, Zhang Y-C, Zhou T. Vital nodes identification in complex networks. Phys Rep 2016;650:1–63.
[14] Hooshmand F, Mirarabrazi F, MirHassani S. Efficient benders decomposition for distance-based critical node detection problem. Omega 2020;93:102037.
[15] Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, et al. Identification of influential spreaders in complex networks. Nat Phys 2010;6(11):888–93.
[16] Freeman LC. A set of measures of centrality based on betweenness. Sociometry 1977;35–41.
[17] Das K, Samanta S, Pal M. Study on centrality measures in social networks: a survey. Soc Netw Anal Min 2018;8:1–11.
[18] Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021;5(3):337–48.
[19] Kahr M, Leitner M, Ruthmair M, Sinnl M. Benders decomposition for competitive influence maximization in (social) networks. Omega 2021;100:102264.
[20] Samadi M, Nagi R, Semenov A, Nikolaev A. Seed activation scheduling for influence maximization in social networks. Omega 2018;77:96–114.
[21] Zhao J, Wen T, Jahanshahi H, Cheong KH. The random walk-based gravity model to identify influential nodes in complex networks. Inform Sci 2022;609:1706–20.
[22] Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. Comput Netw ISDN Syst 1998;30(1–7):107–17.
[23] Kleinberg JM. Authoritative sources in a hyperlinked environment. J ACM 1999;46(5):604–32.
[24] Wen T, Cheong KH. The fractal dimension of complex networks: A review. Inf Fusion 2021;73:87–102.
[25] Wen T, Deng Y. Identification of influencers in complex networks by local information dimensionality. Inform Sci 2020;512:549–62.
[26] Edwards W. The theory of decision making. Psychol Bull 1954;51(4):380.
[27] Wu X, Liao H. Value-driven preference disaggregation analysis for uncertain preference information. Omega 2023;115:102793.
[28] Dias LC, Rocha H. A stochastic method for exploiting outranking relations in multicriteria choice problems. Ann Oper Res 2023;321(1–2):165–89.
[29] Wu X, Liao H. A compensatory value function for modeling risk tolerance and criteria interactions in preference disaggregation. Omega 2023;117:102836.
[30] Dias LC, Cunha MC, Watkins E, Triantaphyllidis G. A multi-criteria assessment of policies to achieve the objectives of the EU marine litter strategy. Mar Pollut Bull 2022;180:113803.
[31] Liu F, Liao H, Al-Barakati A. Physician selection based on user-generated content considering interactive criteria and risk preferences of patients. Omega 2023;115:102784.
[32] Oliveira GD, Dias LC. The potential learning effect of a MCDA approach on consumer preferences for alternative fuel vehicles. Ann Oper Res 2020;293(2):767–87.
[33] Garrido S, Rodríguez-Antón J, Amado CA, Santos S, Dias LC. Development and application of a composite circularity index. J Environ Manag 2023;337:117752.
[34] Dempster AP. Upper and lower probabilities induced by a multivalued mapping. In: Classic works of the Dempster-Shafer theory of belief functions. Springer; 2008, p. 57–72.
[35] Yang J-B, Singh MG. An evidential reasoning approach for multiple-attribute decision making with uncertainty. IEEE Trans Syst Man Cybern 1994;24(1):1–18.
[36] Yang J-B, Xu D-L. Evidential reasoning rule for evidence combination. Artificial Intelligence 2013;205:1–29.
[37] Yang J-B, Xu D-L, Xu X, Fu C. Likelihood analysis of imperfect data. IEEE Trans Syst Man Cybern: Syst 2023.
[38] Liao H, Fang R, Yang J-B, Xu D-L. A linguistic belief-based evidential reasoning approach and its application in aiding lung cancer diagnosis. Knowl-Based Syst 2022;253:109559.
[39] Zhao J, Deng Y. Complex network modeling of evidence theory. IEEE Trans Fuzzy Syst 2020;29(11):3470–80.
[40] Ni L, Chen Y-w, de Brujin O. Towards understanding socially influenced vaccination decision making: An integrated model of multiple criteria belief modelling and social network analysis. European J Oper Res 2021;293(1):276–89.
[41] Zhao J, Song Y, Deng Y. A novel model to identify the influential nodes: Evidence theory centrality. IEEE Access 2020;8:46773–80.
[42] Zhong H, Pajouh FM, Prokopyev OA. Finding influential groups in networked systems: The most degree-central clique problem. Omega 2021;101:102262.
[43] Chen D, Lü L, Shang M-S, Zhang Y-C, Zhou T. Identifying influential nodes in complex networks. Physica A 2012;391(4):1777–87.
[44] Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci 2005;102(46):16569–72.
[45] Hirsch JE. Does the h index have predictive power? Proc Natl Acad Sci 2007;104(49):19193–8.
[46] Hage P, Harary F. Eccentricity and centrality in networks. Social Networks 1995;17(1):57–63.
[47] Freeman LC. Centrality in social networks: Conceptual clarification. Soc Netw 1979;1:215–39.
[48] Latora V, Marchiori M. Efficient behavior of small-world networks. Phys Rev Lett 2001;87(19):198701.
[49] Estrada E, Higham DJ, Hatano N. Communicability betweenness in complex networks. Physica A 2009;388(5):764–74.
[50] Newman ME. A measure of betweenness centrality based on random walks. Social Networks 2005;27(1):39–54.
[51] Li Z, Ren T, Ma X, Liu S, Zhang Y, Zhou T. Identifying influential spreaders by gravity model. Sci Rep 2019;9(1):8387.
[52] Wittenbaum GM, Hubbell AP, Zuckerman C. Mutual enhancement: Toward an understanding of the collective preference for shared information. J Personal Soc Psychol 1999;77(5):967.
[53] Bonacich P. Some unique properties of eigenvector centrality. Social Networks 2007;29(4):555–64.
[54] Martin T, Zhang X, Newman ME. Localization and centrality in networks. Phys Rev E 2014;90(5):052808.
[55] Bonacich P, Lloyd P. Eigenvector-like measures of centrality for asymmetric relations. Social Networks 2001;23(3):191–201.
[56] Wang L, Ma L, Wang C, Xie N-g, Koh JM, Cheong KH. Identifying influential spreaders in social networks through discrete moth-flame optimization. IEEE Trans Evol Comput 2021;25(6):1091–102.
[57] Liu Y, Wang X, Kurths J. Framework of evolutionary algorithm for investigation of influential nodes in complex networks. IEEE Trans Evol Comput 2019;23(6):1049–63.
[58] Fan C, Zeng L, Sun Y, Liu Y-Y. Finding key players in complex networks through deep reinforcement learning. Nat Mach Intell 2020;2(6):317–24.
[59] Morone F, Makse HA. Influence maximization in complex networks through optimal percolation. Nature 2015;524(7563):65–8.
[60] Ren X-L, Gleinig N, Helbing D, Antulov-Fantulin N. Generalized network dismantling. Proc Natl Acad Sci 2019;116(14):6554–9.
[61] Bian T, Hu J, Deng Y. Identifying influential nodes in complex networks based on AHP. Physica A 2017;479:422–36.
[62] Zhang H, Zhong S, Deng Y, Cheong KH. LFIC: Identifying influential nodes in complex networks by local fuzzy information centrality. IEEE Trans Fuzzy Syst 2021;30(8):3284–96.
[63] Zhang Z-K, Liu C, Zhan X-X, Lu X, Zhang C-X, Zhang Y-C. Dynamics of information diffusion and its applications on complex networks. Phys Rep 2016;651:1–34.
[64] Granovetter M. Threshold models of collective behavior. Am J Sociol 1978;83(6):1420–43.
[65] Centola D, Eguíluz VM, Macy MW. Cascade dynamics of complex propagation. Physica A 2007;374(1):449–56.
[66] Gleeson JP, Cahalane DJ. Seed size strongly affects cascades on random networks. Phys Rev E 2007;75(5):056103.
[67] Watts DJ. A simple model of global cascades on random networks. Proc Natl Acad Sci 2002;99(9):5766–71.
[68] Krapivsky PL, Redner S, Volovik D. Reinforcement-driven spread of innovations and fads. J Stat Mech Theory Exp 2011;2011(12):P12003.
[69] Gruhl D, Guha R, Liben-Nowell D, Tomkins A. Information diffusion through blogspace. In: Proceedings of the 13th international conference on world wide web. 2004, p. 491–501.
[70] Trpevski D, Tang WK, Kocarev L. Model for rumor spreading over networks. Phys Rev E 2010;81(5):056102.
[71] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc R Soc Lond Ser A 1927;115(772):700–21.
[72] Xie X, Zhan X, Zhang Z, Liu C. Vital node identification in hypergraphs via gravity model. Chaos 2023;33(1):013104.
[73] Yang J-B, Xu D-L. On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty. IEEE Trans Syst Man Cybern A 2002;32(3):289–304.

[74] Zhou M, Chen Y-W, Liu X-B, Cheng B-Y, Yang J-B. Weight assignment method for multiple attribute decision making with dissimilarity and conflict of belief distributions. Comput Ind Eng 2020;147:106648.
[75] Gleiser PM, Danon L. Community structure in jazz. Adv Complex Syst 2003;6(04):565–73.
[76] Newman ME. Finding community structure in networks using the eigenvectors of matrices. Phys Rev E 2006;74(3):036104.
[77] Adamic LA, Glance N. The political blogosphere and the 2004 US election: divided they blog. In: Proceedings of the 3rd international workshop on link discovery. 2005, p. 36–43.
[78] Hu H-B, Wang X-F. Unified index to quantifying heterogeneity of complex networks. Physica A 2008;387(14):3769–80.
[79] Hébert-Dufresne L, Allard A, Young J-G, Dubé LJ. Global efficiency of local immunization on complex networks. Sci Rep 2013;3(1):1–8.