Contents lists available at ScienceDirect

Journal of King Saud University - Computer and Information Sciences

journal homepage: www.sciencedirect.com

WSLC: Weighted semi-local centrality to identify influential nodes in complex networks

Xiaofeng Wang [a,b,*], Marini Othman [b], Deshinta Arrova Dewi [b], Yonghong Wang [a,b]

[a] Department of Computer Science, Xinzhou Normal University, Xinzhou, Shanxi 034000, China
[b] Faculty of Data Science and Information Technology, INTI International University, Nilai, Negeri Sembilan 71800, Malaysia

ARTICLE INFO

Keywords:
Complex networks
Influential nodes
Weighted semi-local centrality
Extended neighborhood concept

ABSTRACT

Identifying and ranking influential nodes in complex networks is a critical aspect to study the survival and robustness of networks. Many ongoing researches have proposed centrality metrics to address this problem, so that the performance of each is attributed to specific scenarios. For example, metrics based on local structure have low ranking accuracy due to the use of limited information, and metrics based on global structure suffer from high complexity. Meanwhile, metrics based on semi-local structure are amazingly well, but an efficient centrality for identifying influential nodes is still not available due to differences in the structure and scale of networks. In addition, most semi-local centrality metrics only consider one aspect of each node's information, and their development still faces serious challenges. This paper develops a Weighted Semi-Local Centrality (WSLC) to identify influential nodes in complex networks based on extended neighborhood concept. Here, several different weights are investigated to find the best performance on WSLC. We use the extended neighborhood concept to select the nearest neighbors, which considers the global information of the network in a limited and efficient way to calculate the ranks. Here, a distributed approach is presented that can cut a subgraph of the entire network for each node with low complexity. This subgraph contains neighbors with different hops, which are used to maintain high efficiency when facing large-scale networks. In addition to the importance of the node itself, WSLC also combines the importance of the node's nearest neighbors with different hops for ranking. Therefore, defining semi-local structure with a distributed approach as well as using an efficient edge weighting policy differentiates WSLC from other existing centrality metrics. The evaluation of WSLC has been done through several real-world networks using Kendall's correlation. The effectiveness of WSLC under the SIR infection spreading model has been verified by extensive simulations compared to state-of-the-art centrality metrics.

1. Introduction

In recent decades, the emergence of social relationships has moved towards becoming more complex and forming institutions with complex relations. Nowadays, these institutions are considered as complex networks that have some common properties such as small-world, clustering coefficient, and scale-independent (Berahmand et al., 2022). Every complex network contains some nodes and complex relationships between them. The heterogeneity of the structure of complex networks leads to the appearance of different roles of nodes (Zhao et al., 2023d, 2023e). Hence, some nodes can affect the performance and structure of the network to a greater extent. These nodes are known as seed/influential nodes in complex networks and have significant effectiveness in controlling or diffusion of information.

In the process of spreading information, network structure and relationships between users are very important. Also, spreading mechanisms are influenced by a small group of users. Therefore, the selection of primary nodes for spreading a specific behavior can be different (Zhao et al., 2022). In order to achieve the maximum influence, a small group of nodes can be selected and influence the opinion of other nodes through them. Identifying such nodes is one of the challenging topics in the field of network analysis (Li et al., 2019). The problem of identifying influential nodes in complex networks has great theoretical and practical importance, which has been understood by the research society (Cao et al., 2023; Wang et al., 2023). In general, the influential nodes in the field of network information mining are known as an open problem

* Corresponding author at: Department of Computer Science, Xinzhou Normal University, Xinzhou, Shanxi 034000, China.
E-mail addresses: xz103089@126.com (X. Wang), marini.othman@newinti.edu.my (M. Othman), deshinta.ad@newinti.edu.my (D.A. Dewi), yonghong.wang.tch@gmail.com (Y. Wang).

https://doi.org/10.1016/j.jksuci.2023.101906
Received 3 November 2023; Received in revised form 26 December 2023; Accepted 26 December 2023
Available online 27 December 2023

because it has many applications in the real world. The most important of these applications include information propagation, online advertising, rumor control, time series, marketing, advertising, and opinion monitoring (Rezaeipanah et al., 2020).

Effective node identification approaches are divided into two general categories: activity-based techniques and centrality-based techniques (Salavati et al., 2019). Activity-based techniques such as heat diffusion and similarity measure depend on the type of activity of nodes in a network (Mohammadian et al., 2022; Zhang et al., 2022). Centrality-based techniques such as degree, closeness, and betweenness are defined based on the unique characteristics of each node in a network (Berahmand et al., 2018). In the last few decades, many centrality metrics have been proposed considering the topological structure and dynamics of complex networks (Jannesari et al., 2023; Xue et al., 2023). In general, each centrality metric measures the rank of each node in the entire network as influential. For example, the degree centrality metric shows the rank of each node with that node's degree, so that the importance of a node with a higher degree is greater than other nodes.

Considering the neighborhood level, centrality metrics are divided into local, semi-local, and global categories (Yue et al., 2023; Zhao et al., 2023a). Local metrics only use the information of first-level neighbors, while global metrics consider the information of the entire network to calculate influence. Meanwhile, semi-local metrics use the information of neighbors with different levels to compromise between complexity and performance (Rostami et al., 2023). Basically, semi-local centralities are defined based on fixed-length nearest neighbor information and perform better for large-scale networks (Yang et al., 2022). However, most of the centrality metrics in the semi-local category only use the information of neighbors of the first and second levels to estimate influence and do not apply topological connections.

In recent years, several centrality metrics have been proposed that use the information of neighbors with higher levels (Liu et al., 2016; Zhang et al., 2023). However, the information of neighbors with higher levels to calculate influence improves the accuracy but increases the complexity. Therefore, approaches to extract more information from the network with low complexity can overcome this problem. Considering the extended neighborhood concept, this paper tries to present a distributed approach for extracting nearest neighbors' information with different levels. Here, each node can independently identify a subgraph of the network with nearest neighbors. This subgraph is used to calculate the influence of each node, which can significantly reduce the complexity. Moreover, the insight in the literature shows that edge information is also involved in centrality. Hence, in addition to node information, we use edge information to identify influential nodes. Here, the edges in the extracted subgraph are weighted and the weight information is applied to calculate influence. Since how the weights are calculated is so important, we will examine several different weights. According to the stated concepts, this paper proposes a Weighted Semi-Local Centrality (WSLC) based on extended neighborhood concept to identify influential nodes in complex networks.

The main contribution of this paper is as follows:

- Each node independently finds the nearest neighbors with different levels based on the extended neighborhood concept. The nearest neighbors are extracted as a subgraph from the network and applied to calculate the influence.
- In addition to node information, edge information is used to calculate the influence. The edge information contains an edge weight assignment policy that is applied to the extracted subgraph.
- A weighted semi-local centrality based on the extended neighborhood concept is developed, which simultaneously uses the information of the node itself and the nearest neighbors to calculate the influence.

The following is the structure of this paper. The research literature and related works for identifying the influential nodes is given in Section 2. Section 3 focuses into proposed WSLC centrality metric in depth. The experimental setup, along with the test outcomes are specified in Section 4. Finally, Section 5 concludes the paper.

2. Literature review

In this section, we describe some well-known centrality metrics and then review some state-of-the-art centralities from the semi-local category.

2.1. Preliminaries

A complex network can be imagined with a graph $G = (V, E)$, where $v \in V$ is the set of nodes and $e_{u,v} \in E$ is the set of edges. Let $e_{u,v}$ be the link between nodes u and v in an undirected network such as G. According to this definition, $N = |V|$ is the total number of network nodes, and $M = |E|$ is the total number of network edges. Meanwhile, consider $G = (V, E, W)$ as a graph for a weighted network, where W is the set of edge weights. Here, $w_{u,v} \in W$ represents the weight associated with the edge $e_{u,v}$.

The neighborhood of the first level in the graph G is defined by the adjacency matrix A, where each $a_{u,v} \in A$ represents the connection status between nodes u and v. For example, $a_{u,v} = 1$ indicates a link between nodes u and v, and $a_{u,v} = 0$ indicates the absence of any link. Also, $\Gamma(v)$ is defined as the set of all neighbors of the first level with node v, and $k_v = |\Gamma(v)|$ refers to the number of these neighbors. In addition, the number of hops (or distance) between nodes u and v via the shortest path is defined by $\delta_{u,v}$.

2.2. Centrality metrics

To identify influential nodes in different networks, many techniques have been introduced so far (Guo et al., 2023; Huang et al., 2023). Each of these methods have been developed by considering different aspects of the network structure, such as the type of communication, the type of target, and the characteristics of the network. These methods are known as centrality metrics in complex networks. According to the network information used to measure the rank of nodes, centrality metrics are divided into three general categories: local centralities, semi-local centralities and global centralities (Cao et al., 2022; Zhao et al., 2023b). As shown in Table 1, various centrality metrics of each category have been proposed so far.

Local centralities ignore the global structure of the topology and are often less relevant to real-world networks. However, these metrics are simple and have little complexity because they only apply the degree of a node's neighbors to calculate its rank (Forouzandeh et al., 2021). Global centralities provide better results than local centralities, because

Table 1
Different types of centrality metrics.

Local centralities	Semi-local centralities	Global centralities
Degree (Freeman, 2002)	NCvoteRankcentrality (Kumar and Panda, 2020)	Betweenness (Freeman, 1977)
Cluster coefficient (Serrano and Boguna, 2006)	k-shell (Kitsak et al., 2010)	Closeness (Sabidussi, 1966)
PageRank (Brin and Page, 1998)	k-shell decomposition (Sheng et al., 2020b)	Eigenvector (Bonacich, 2007)
Trust-PageRank (Sheng et al., 2020b)	Semi-local centrality (Chen et al., 2012)	Relative change of average shortest path (Lv et al., 2019)
Local neighbor contribution (Dai et al., 2019)	Mixed degree decomposition (Zeng and Zhang, 2013)	Global and local structure (Sheng et al., 2020a)
Normalized local centrality (Zhao et al., 2018)	Local structural centrality (Gao et al., 2014)	Global importance of a node (Zhao et al., 2020)

they use the information of holistic network for ranking. However, these metrics are inefficient for large-scale networks due to high time complexity (Fan et al., 2020).

In recent years, the scale of online social platforms is growing to billions of users. Hence, local centralities will be unusable due to limited information utilization and global centralities due to high complexity. Recently, some metrics focus on mixed local and global structures as semi-local centralities (Yang et al., 2020). Semi-local centralities simultaneously consider first-level neighbors and next-nearest neighbors to measure influence. Chen et al. (2012) showed that these metrics provide a good ability to rank nodes by balancing accuracy and complexity.

One of the most famous local centrality metrics is Degree Centrality (DC), which considers the degree of the node as influence (Freeman, 2002). PageRank (PR) and Trust-PageRank (TPR) are other local centrality metrics. By focusing on the ranking of web pages, PR can calculate the influence of nodes in complex networks (Brin and Page, 1998). TPR is the same as PR except that trust is applied to neighbors while ranking nodes (Sheng et al., 2020b). Local centrality metrics are simple and fast, but have low accuracy due to access to limited information.

Betweenness Centrality (BC), Closeness Centrality (CC), and Eigenvector Centrality (EC) are among the most famous global centrality metrics. BC considers the number of observations of a node in all the shortest paths in the network for its influence (Freeman, 1977). CC considers the lowest average distance to other nodes as the influence of a node (Sabidussi, 1966). EC measures the influence of a node in the network by the normalized eigenvector belonging to the largest eigenvalue (Bonacich, 2007). Lv et al. (2019) proposed a global centrality metric based on Average Shortest Path (ASP) theory, known as Relative change of ASP (RASP). RASP includes changes in the average shortest path after removing a node from the network. Global centrality metrics are highly accurate, but suffer from high complexity due to the use of information from the entire network.

Semi-local centrality metrics have attracted more attention to identify influential nodes in complex networks because they strike a balance between complexity and accuracy. Semi-local Centrality (SC) and Mixed Degree Decomposition (MDD) are the most common metrics of semi-local centrality. SC calculates the influence of a node by simultaneously considering the degree of neighbors in the first and second levels (Chen et al., 2012). MDD uses K-Shell (KS) index to identify influential nodes, where exhausted degree and residual degree are simultaneously applied (Zeng and Zhang, 2013).

Table 2 summarizes more details of the investigated metrics.

2.3. Related works

So far, many centrality metrics have been devised to find influential nodes in complex networks (Kang et al., 2016; Ullah et al., 2022; Zhang et al., 2023). Each of these metrics has some shortcomings and own points. In fact, the type of influence in a network does not appear as a "natural" concept and can be different from one network to another. Therefore, each of the centrality metrics may interpret influence to rank nodes with different viewpoints. For example, the BC metric defines influence as an index of bridging between nodes, whereas the CC metric highlights the minimum distance to connect to other nodes (Zhou et al., 2021; Wu et al., 2023). This shows that in the analysis of complex networks, "semantic profiles" are different from centrality methods.

In recent years, semi-local centrality metrics have received the attention of the research society due to the balance between accuracy and complexity (Torabi et al., 2022). These metrics consider both local and global information from the network structure and use multiple characteristics of nodes to measure influence as much as possible (Ni et al., 2021; Wang et al., 2022). Due to the large amount of literature on semi-local centrality metrics, we limit this section to reviewing only these works. A summary of the investigated semi-local centralities along with their formula is given in Table 3. For a better understanding of the proposed centrality metric, we have also included the details of the WSLC in this table.

The Degree and Importance of Lines (DIL) centrality semi-local metric was proposed by Liu et al. (2016) where the influence of nodes and edges is applied in the influence calculation. DIL calculates the importance of edges by considering the characteristics of the nodes that are linked to them. The authors measure the weight of each edge in DIL based on fungibility and connectivity. This metric identifies bridge nodes with high accuracy and has an acceptable complexity for processing large-scale networks. However, DIL is inefficient for identifying influential nodes with the same degrees.

Kang et al. (2016) proposed a Weighted Semi-Local Centrality Criterion (WSLCC) to identify influential nodes in complex networks. WSLCC tries to reflect the violation of local centrality metrics by simultaneously considering semi-local information and weighted degree as the influence strength of the node. WSLCC is inspired by evidence

Table 2
Details of centrality metrics.

Reference	Metric	Category	Formula	Description of parameters
Freeman (2002)	DC	Local	$DC(v) = k_v$	–
Brin and Page (1998)	PR	Local	$PR(v) = \frac{1-\alpha}{N} + \alpha\sum_{u\in\Gamma(v)}\frac{PR(u)}{k_v}$	α is the jump probability.
Sheng et al., (2020b)	TPR	Local	$TPR(v) = \frac{1-\alpha}{N} + \alpha\sum_{u\in\Gamma(v)}T(u,v).$ $TPR(u)$	$T(u,v)$ is the trust value between nodes u and v.
Freeman (1977)	BC	Global	$BC(v) = \sum_{u\neq w\neq v\in V}\frac{\delta_{u,w}(v)}{\delta_{u,w}}$	–
Sabidussi (1966)	CC	Global	$CC(v) = \frac{1}{\sum_{u\neq v\in V}\delta_{u,v}}$	–
Bonacich (2007)	EC	Global	$EC(v) = \mu\sum_{u\in\Gamma(v)}x_u$	μ is a constant value based on the largest eigenvalue of A, and x_u is the influence of node u according to the normalized eigenvector belonging to the largest eigenvalue of A. Let $x = [x_1, x_2, \cdots, x_{\lvert V\rvert}]^T$ be an eigenvector associated with the eigenvalue μ^{-1} of A.
Lv et al. (2019)	RASP	Global	$RASP(v) = \frac{\lvert ASP[G] - ASP[G_v]\rvert}{ASP[G]}$	$ASP[G]$ is the average number of steps along with shortest paths for all possible pairs of nodes in G. Also, G_v is the network G after node v is removed.
Chen et al. (2012)	SC	Semi-local	$SC(v) = \sum_{u\in\Gamma(v)}\sum_{w\in\Gamma(u)}k_w$	–
Zeng and Zhang (2013)	MDD	Semi-local	$MDD(v) = K_r + \lambda.K_e$	λ is a tunable balance parameter, K_r is the residual degree, and K_e is the exhausted degree.

Table 3
Summary of recent semi-local centralities.

Reference	Metric	Formula	Description of parameters
Kang et al. (2016)	WSLCC	$WSLCC(v) = \sum_{u\in\Gamma(v)}\sum_{w\in\Gamma(u)}(N^w(w) + k'_w) + k'_v$	$N^w(w)$ is the sum of the weighted degree of node w as well as its 2-hop and 3-hop neighbors. Also, $k'_v = \sqrt{k_v\sum_{u\in\Gamma(v)} w_{u,v}}$.
Shao et al. (2019)	NL	$NL(v) = \sum_{u\in\Gamma(v)}\left(\varphi(u) + I_{a_{u,v}}\cdot\frac{k_v - 1}{k_v + k_u - 2}\right)$	$\varphi(u)$ is the number of neighbors with path-length equal to 2 of u.
Ullah et al. (2021)	LGC	$LGC(v) = \frac{k_v}{N}\cdot\sum_{u\in\Gamma(v)}\frac{\sqrt{k_u + \alpha}}{\delta_{u,v}}$	α is a tunable parameter that controls the effect of degree.
Ullah et al. (2022)	EVC+	$EVC^{+}(v) = \sum_{u\in\Gamma(v)}\sqrt{\frac{2\alpha(k_v(ks_v + ks_u))}{\delta_{u,v}}}$	ks_v is KS of node v, and α is a tunable parameter to control the degree effect.
Zhang et al. (2023)	INASP	$INASP(v) = \frac{a_1.k_v + a_2.\sum_{l=1}^{L}\sum_{u\in V^{+l}(v)}\frac{k_u}{l}}{L.N_v} + a_3.\widehat{ASP}[\widehat{G}_v]$	L is the maximum hop for the neighborhood, $V^{+l}(v)$ is the neighbors of node v in the l-hop, and $\widehat{ASP}[\widehat{G}_v]$ is the LRASP index for node v. The coefficients of each section are defined by parameters a_1, a_2, and a_3.
Proposed metric	WSLC	$WSLC(v) = a_1.I_{Node}(v) + a_2.I_{Local}(v) + a_3.I_{Semi\_Local}(v)$	I_{Node} is the importance of the node itself, I_{Local} is the importance of the local node, and I_{Semi_Local} is the importance of the semi-local node. Also, a_i are the tunable influence coefficients.

theory and semi-local centrality presented by Gao et al. (2013), with the difference that WSLCC is developed on a weighted network and considers the connections of multiple layer neighbors.

Shao et al. (2019) presented a semi-local centrality method based on DIL based on Neighbors and the importance of Links (NL). NL applies the importance of all second-level neighbors to rank nodes. In addition, the authors use the unsubstitutability and connectivity of edges to apply the topological position of nodes in influencing estimation. Therefore, NL considers both topological position and semi-local structure for ranking.

Ullah et al. (2021) addressed the disadvantages of local and global metrics by simultaneously considering local and global topological aspects. The authors developed the Local-and-Global-Centrality (LGC) metric, which includes three definitions: local-influence, global-influence, and total influence. In local-influence, the ratio of node degree to all network nodes is considered. Global-influence includes the importance of neighboring nodes as well as shortest distances. Finally, LGC is a combination of local-influence and global-influence definitions.

Local RASP (LRASP) as a semi-local centrality metric was presented by Hajarathaiah et al. (2022). LRASP considers part of the network to calculate RASP. Here, all neighbors of a node up to a fixed level are cut as an induced subgraph of the entire network and considered as input for the RASP metric. The results show that LRASP improves the balance between complexity and accuracy compared to RASP.

Ullah et al. (2022) used the escape velocity formula to identify influential nodes and introduced the Escape Velocity Centrality (EVC) metric. EVC considers both local and global information and measures the rank of each node by combining shortest distance and degree. Since degree alone is not able to show the influence of nodes, the authors proposed EVC+ as an extended version of EVC. EVC+ increases performance by simultaneously including degree and KS in EVC.

A semi-local centrality metric combining the LRASP index, the importance of the node itself, and the importance of the node's neighbors was proposed by Zhang et al. (2023). Let INASP be the symbol to denote this metric. The importance of the node itself in INASP is applied by degree. The influence of the nearest neighbors is measured using the influence of the connected nodes and the hop-count between them. Here, the nearest neighbors are defined by first, second and third level neighbors. Here, instead of the entire network, a small subgraph is extracted for use by LRASP.

Weighted Hybrid Centrality (WHC) is another semi-local metric that applies information from both nodes and edges to calculate influence (Shetty and Bhattacharjee, 2022). Since the importance of all edges is considered the same in unweighted networks, WHC develops an edge weighting approach to apply the frequency of interactions between each pair of nodes during ranking. In addition, WHC includes several well-known centrality methods such as degree, KS, and EC.

3. Proposed centrality metric

Semi-local centralities are more effective compared to local centralities such as degree and PageRank, as well as global centralities such as betweenness and closeness in dealing with large-scale networks. Recent works have shown that centrality in a semi-local structure depends not only on the node itself but also on its nearest neighbors (Hajarathaiah et al., 2022). However, most local centrality metrics are defined only based on the number of first- and second-level neighbors and ignore the topological connections between neighbors (Masdari et al., 2020). Meanwhile, metrics that consider both topological connectivity and the number of neighbors to identify influential nodes are still under development (Zhao et al., 2023c). Topological communication refers to the position of nodes and connections between them in the network, which can be applied by considering neighborhood information at different levels. Considering the entire network structure requires considering all levels of neighborhood and is inefficient for large-scale networks. Therefore, neighborhood levels should be applied in a limited way in ranking nodes. Also, the identification of neighbors with different levels should be done with low computational complexity to maintain performance in the face of large-scale networks.

Insights in the literature show that the influence of a node depends not only on itself but also on its nearest neighbors. Also, the importance of edges in measuring influence should not be assumed to be equal, as this is an unrealistic assumption of social interactions. With this motivation, we propose WSLC as a weighted centrality metric based on the extended neighborhood concept. WSLC includes three features to measure the influence of each node: 1) the importance of the node itself in the network (i.e., Node-Influence), 2) the importance of that node's direct neighbors in the network (i.e., Local-Influence), and 3) the importance of the nearest neighbors of that node with different levels in the network (i.e., Semi-Local-Influence). The combination of these three features by WSLC can apply network structure, connections between neighbors and topology heterogeneity to rank nodes.

Let $I_{Node}(v)$ denote the importance of node v in terms of Node-Influence. Also, let $I_{Local}(v)$ and $I_{Semi\_Local}(v)$ be the importance of node v in terms of Local-Influence and Semi-Local-Influence, respectively. By combining these three features, WSLC calculates the total influence of node v in the network, as defined by Eq. (1).

$$WSLC(v) = a_1.I_{Node}(v) + a_2.I_{Local}(v) + a_3.I_{Semi\_Local}(v) \quad (1)$$

where a_1, a_2 and a_3 are tunable influence coefficients for Node-

Influence, Local-Influence and Semi-Local-Influence features, respectively.

All three features defined in WSLC are normalized between 0 and 1 before combining so that the effect of all of them is the same in the total influence measurement. The details of the three features used in WSLC are as follows.

3.1. Node-influence

The degree of a node is one of the most important factors to determine its centrality in complex networks. Hence, we define Node-Influence in WSLC based on the node degree. However, the density of the network can affect the degree of the node as a centrality factor. Suppose the degree of node v is equal to x. Obviously, the importance of node v with degree x is higher in a network with low connections than in a network with high connections. Hence, we define Node-Influence for node v based on both degree and density, as shown in Eq. (2).

$$I_{Node}(v) = \frac{k_v}{k_{max} + D_G} \tag{2}$$

where k_{max} is the largest degree of the network. Also, D_G denotes the density for network G, which is defined by Eq. (3) for undirected simple networks.

$$D_G = \frac{2M}{N(N-1)} \tag{3}$$

3.2. Local-influence

The insight in the literature shows that the greater the influence of a node compared to its neighbors, the more likely that node will be influential in the network. However, the number of neighbors should not be neglected because it is directly proportional to the influence of the node. Therefore, the importance of a node depends not only on itself but also on its neighbors. On the other hand, the contribution of each neighbor in the influence measurement should not be neglected. According to these definitions, we calculate $I_{Local}(v)$ by Eq. (4).

$$I_{Local}(v) = \frac{1}{k_v}.\sum_{u\in\Gamma(v)} \frac{\sqrt{w_{u,v}.k_v}}{k_u + k_v} \tag{4}$$

where $w_{u,v}$ is the weight associated with the edge $e_{u,v}$, which represents the connection contribution between nodes v and u.

The research society has strong evidence of the significant impact of the heterogeneity of connection structures and diverse patterns on the understanding of influence dynamics in complex networks. Therefore, heterogeneity of topologies is a crucial aspect of centrality. Although some state-of-the-art metrics have been developed considering the connection strength between nodes as weights, recent studies show that there is no reliable and consistent metric in facing different networks. This weakness has been reduced in recent works by combining different centrality metrics and simultaneously considering the weight of the edges. However, most of these works have ignored the heterogeneity of connectivity structures in complex networks. With this motivation, we apply the effectiveness of edges through a weight in the penetration measure. Since the influence of a node depends on how the weight is calculated, we introduce several different weight policies. In the following, six weighting policies are defined to calculate $w_{u,v}$ as the weight between nodes v and u.

- *Common Neighbors (CN):* CN refers to the number of common nodes between two nodes, as defined in Eq. (5) (Lorrain and White, 1971).

$$CN_{u,v} = |\Gamma(u) \cap \Gamma(v)| \tag{5}$$

- *Jaccard Coefficient (JC):* JC refers to the number of common neighbors relative to the total number of neighbors between two nodes, as defined in Eq. (6) (Jaccard, 1901).

$$JC_{u,v} = \frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|} \tag{6}$$

- *Average Degree (AD):* This policy is defined based on the average degree of nodes u and v, as shown in Eq. (7).

$$AD_{u,v} = \frac{k_u + k_v}{2} \tag{7}$$

- *Neighbors Degree (ND):* This policy defines the weight between nodes u and v based on the sum of the average degrees of the neighbors of each of these nodes, as shown in Eq. (8).

$$ND_{u,v} = \frac{\sum_{w\in\Gamma(u)} k_w}{k_u} + \frac{\sum_{w\in\Gamma(v)} k_w}{k_v} \tag{8}$$

- *Reputation-Optimism (RO):* The weighting policy of RO includes the factors 'Reputation' and 'Optimism', which refer to the popularity of a user in the network and the following of more users, respectively. Let the weighted RO policy be calculated through Eq. (9).

$$RO_{u,v} = \frac{2.k_u.k_v}{k_u + k_v} \tag{9}$$

- *Katz Index (KI):* KI contains all paths with a given maximum length between two nodes. This policy is calculated based on the path frequency and the path length factor, as defined in Eq. (10) (Katz, 1953).

$$KI_{u,v} = \sum_{l=1}^{L} \beta^l.\left|\mathscr{P}_{i,j}^{<l>}\right| \tag{10}$$

where $\left|\mathscr{P}_{i,j}^{<l>}\right|$ represents the number of paths between u and v with l-hop, and β is a damping coefficient to reduce the effect of long paths. Setting β to 0.05 is recommended by researchers (Shao et al., 2019). It is worth noting that $\left|\mathscr{P}_{i,j}^{<l>}\right|$ can be calculated by raising the adjacency matrix A to the power of l. Also, L is the maximum neighborhood level, which is always set equal to 2 for the Local-Influence feature.

3.3. Semi-local-influence

The analysis of complex networks shows that the information of the nearest neighbors is effective in the ranking of nodes. However, different definitions of nearest neighbors have been proposed by researchers. For example, Chen et al. (2012) uses first-degree neighbors to calculate influence. Zhang et al. (2023) proved that the nearest neighbors of a node with fixed and limited hops contain more effective information for calculating influence. Obviously, considering all neighbor levels results in a global centrality metric, which is inefficient for large-scale networks. Accordingly, WSLC limits the neighbor levels to a maximum of L so that the information of 1-hop, 2-hop, …, and L-hop neighbors can be applied to calculate the influence. Eq. (11) shows how to calculate $I_{Semi\_Local}(v)$ in WSLC.

$$I_{Semi\_Local}(v) = \frac{1}{SZ\left[G_v^{<L>}\right]} \cdot \sum_{l=2}^{L} \left(\beta^l \cdot \sum_{u \in \Gamma_l(v)} \frac{\sqrt{\widehat{w}_{u,v}.k_v}}{\delta_{u,v}.(k_u + k_v)} \right) \tag{11}$$

where $G_v^{<L>}$ is a subgraph of G containing all neighbors up to level L of node v. SZ is a function that returns the total number of nodes in the subgraph $G_v^{<L>}$. $\Gamma_l(v)$ is the set of neighbors for node v with l-hop, and β is a damping coefficient to reduce the effect of long paths. Setting β to 0.05 is recommended by researchers (Gao et al., 2013). Furthermore, $\widehat{w}_{u,v}$ is the weight between nodes u and v with a reliable path.

It is worth noting that Semi-Local-Influence also includes distance to neighbors, as it is inversely proportional to influence. In general, the nodes u and v in $I_{Semi\_Local}(v)$ may be connected through some paths, so the edge contribution must be applied through the path. Here, $\widehat{w}_{u,v}$ contains the weight multiplication of edges between nodes u and v via the shortest available path. Since the weights of the edges depict the strength of the connections, the multiplication of the weights of the edges affects the path length and provides the reliability of the path.

An important challenge in $I_{Semi\_Local}(v)$ is to obtain $\Gamma_l(v)$ in large-scale graphs. The set of neighbors in the first level is available through the adjacency matrix. Obtaining $\Gamma_2(v)$ as neighbors of the second level has negligible complexity, although it depends on the degree of the node v and the neighbors. However, finding $\Gamma_l(v)$ for larger values of l imposes a higher computational complexity on the centrality metric. To address this challenge, WSLC identifies the nearest neighbors at different levels by considering the extended neighborhood concept. Zhang et al. (2023) introduced a distributed strategy based on the extended neighborhood concept, in which a semi-local subgraph of the entire network is extracted for each node. In this strategy, each node independently creates a neighbor table that contains all neighbors up to the maximum level L. According to this idea, we use this semi-local subgraph to obtain $\Gamma_l(v)$.

The identification of the nearest neighbors for a node v is done by considering the extended neighborhood concept through two control packets ‘P1’ and ‘P2’. Node v broadcasts the control packet ‘P1’ with a “Max.HopCount” equal to L in the network. Hence, packet ‘P1’ is received by all neighbors with maximum level L from node v. After that, all the nodes that received the ‘P1’ packet send their properties including NodeID, degree and distance/hop to node v through the ‘P2’ control packet.

Algorithm 1 shows a pseudo-code of the proposed distributed approach based on the extended neighborhood concept to identify near neighbor nodes. We assume that the graph is accessible using an adjacency matrix and each node has the set of all its adjacent edges. Therefore, each node can discover all its neighbors through the adjacency matrix in linear time. Algorithm 1 is executed recursively similar to Depth First Search (DFS) algorithm. The complexity of DFS is $O(N+M)$ with the adjacency matrix available. However, our algorithm imposes a depth limit up to a maximum level of L. Hence, similar to a Depth-Limited Search (DLS) algorithm, Algorithm 1 traverses the graph only up to a maximum depth L. Therefore, the time complexity of our algorithm similar to DLS is $O(b^L)$, where b is the branching coefficient.

Algorithm 1. Pseudocode of identifying neighboring nodes based on extended neighborhood

Input: Maximum neighborhood level L, node v.
Output: $\Gamma_L(v)$ as neighboring nodes v to the maximum level L.

(*continued on next column*)

Algorithm 1. Pseudocode of identifying neighboring nodes based on extended neighborhood (*continued*)

$\Gamma_L(v) \leftarrow$ Add node v to the list of discovered nodes;
if Max.HopCount $(v) = 0$ **then**
  Return $\Gamma_L(v)$;
end
foreach $u \in \Gamma(v)$ **do**
  if $u \notin \Gamma_L(v)$ **then**
    $L = L - 1$;
    Create packet ‘P1’ with “NodeID $= v$”, “SourceID $= u$”, and “Max. HopCount $= L$”;
    Send packet ‘P1’ to node u;
    Create packet ‘P2’ with “NodeID $= u$”, “SourceID $= v$”, and “Information (u)”;
    Send packet ‘P2’ to node v;
    Running Algorithm 1 (L, u, Γ_L) as a recursive algorithm;
  end
end

3.4. *Example of WSLC*

To better understand the proposed centrality metric, we describe a numerical example of WSLC. Consider Fig. 1, where there is a simple unweighted graph with 16 nodes and 18 edges. Here, WSLC is calculated for node 1, assuming $L = 3$, $w_{u,v} = ND_{u,v}$, and $\beta = 0.5$. According to Eq. (2), $I_{Node}(1)$ is calculated as follows:

$$D_G = \frac{2*18}{16*(16-1)} = 0.15$$

$$I_{Node}(1) = \frac{5}{6+0.15} = 0.81$$

According to Eq. (4), $I_{Local}(1)$ is calculated as follows:

$$w_{1,2} = \frac{12}{5} + \frac{5}{1} = 7.4$$

$$w_{1,3} = \frac{12}{5} + \frac{8}{2} = 6.4$$

$$w_{1,5} = \frac{12}{5} + \frac{11}{3} = 6.07$$

$$w_{1,8} = \frac{12}{5} + \frac{14}{3} = 7.07$$

$$w_{1,9} = \frac{12}{5} + \frac{9}{3} = 5.4$$

$$I_{Local}(1) = \frac{1}{5} * \left(\frac{\sqrt{7.4*5}}{1+5} + \frac{\sqrt{6.4*5}}{2+5} + \frac{\sqrt{6.07*5}}{3+5} + \frac{\sqrt{7.07*5}}{3+5} + \frac{\sqrt{5.4*5}}{3+5} \right) = 0.78$$

To calculate $I_{Semi\_Local}(1)$, first the subgraph associated with node 1 is extracted considering the maximum level of neighborhood equal to 3. Let $G_1^{<3>}$ be the extracted subgraph. According to Eq. (11), $I_{Semi\_Local}(1)$ can be calculated as follows:

$$I_{Semi_{Local}}(1) = \frac{1}{16} * \left(0.5^2 * \left[\frac{14.6}{16} + \frac{13.5}{16} + \frac{12.7}{12} + \frac{14.6}{22} \right]_{l=2} + 0.5^3 * \left[\frac{31.2}{18} + \frac{39.4}{18} + \frac{39.4}{18} + \frac{39.4}{18} + \frac{39.4}{18} + \frac{39.4}{18} \right]_{l=3} \right) = 0.26$$

where $\Gamma_2(1)$ includes nodes 4, 6, 10, and 11, and $\Gamma_3(1)$ includes nodes 7, 12, 13, 14, 15, and 16.

Finally, $WSLC(1)$ is calculated according to Eq. (1) as follows:

$$WSLC(1) = (0.25*0.81) + (0.3*0.78) + (0.45*0.26) = 0.5556$$

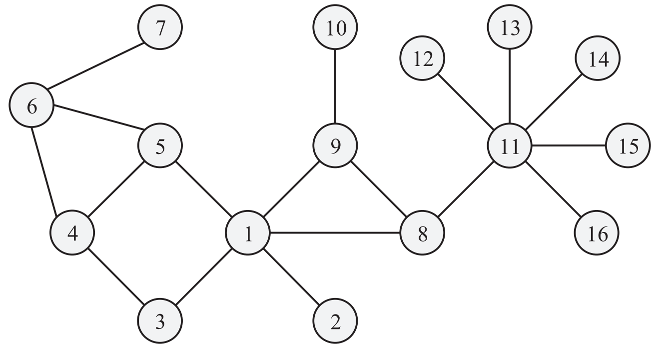

Fig. 1. A simple unweighted graph with 16 nodes and 18 edges.

Here, $\alpha_1 = 0.25$, $\alpha_2 = 0.3$, and $\alpha_3 = 0.45$ are assumed. The WSLC values for other nodes in Fig. 1 are reported in Table 4. For better understanding, the reader can download the MATLAB source code for this example from https://github.com/WSLCCentrality/InfluentialNodes.

3.5. Pseudo-code

Algorithm 2 shows the pseudo-code of proposed WSLC centrality metric which is developed based on the extended neighborhood concept. This algorithm returns the importance of node v in graph G based on the WSLC metric.

Algorithm 2 is designed to calculate the importance of node v in graph G. Lines 1 and 2 define the details of the graph, including the number of nodes, the number of edges, and the adjacency matrix. Line 3 sets the weight of each edge of G. Line 4 calculates the density of the graph G. Line 5 calculates local-influence. Line 6 finds the set of all first-level neighbors of node v based on the adjacency matrix. Line 7 sets $I_{Local}(v)$ to an initial value of zero. Lines 8–11 calculate the value of $I_{Local}(v)$ according to the definition of Local-Influence. Line 12 sets $I_{Semi\_Local}(v)$ to an initial value of zero. The subgraph associated with $G_v^{<L>}$ is created in line 13. Line 14 repeats up to a maximum of L as the neighborhood level. In line 15, the set of all neighbors up to the maximum level l of v according to extended neighborhood concept are identified. Line 16 is for iteration per neighbor of $\Gamma_l(v)$. The shortest path between v and any $u \in \Gamma_l(v)$ on $G_v^{<L>}$ is found in line 17. Lines 18–21 calculate the weights between u and v based on the reliable paths. Line 22 finds the length of the shortest path between u and v. Lines 23 and 25 calculate the value of $I_{Semi\_Local}(v)$ according to the definition of Semi-Local-Influence. Line 25 calculates the final semi-local-influence. Line 28 calculates total-influence. Finally, the algorithm returns $WSLC(v)$ as output in line 29.

Table 4
WSLC metric results for all nodes in Fig. 1.

Nodes	I_{Node}	I_{Local}	I_{Semi_Local}	*WSLC*	Ranks
1	0.8130	0.7806	0.2627	0.5556	1
2	0.1626	0.4534	0.1139	0.2279	10
3	0.3252	0.6207	0.0955	0.3105	8
4	0.4878	0.7555	0.0436	0.3682	5
5	0.4878	0.6556	0.0629	0.3469	6
6	0.4878	0.7842	0.0290	0.3703	4
7	0.1626	0.5774	0.0164	0.2213	11
8	0.4878	0.6154	0.0851	0.3449	7
9	0.4878	0.7877	0.0494	0.3805	3
10	0.1626	0.6124	0.0166	0.2318	9
11	0.9756	0.9008	0.0184	0.5224	2
12	0.1626	0.3869	0.0379	0.1738	12
13	0.1626	0.3869	0.0352	0.1725	13
14	0.1626	0.3869	0.0328	0.1715	14
15	0.1626	0.3869	0.0307	0.1705	15
16	0.1626	0.3869	0.0288	0.1697	16

4. Results and discussions

To evaluate the performance of the proposed centrality metric, we use the SIR model as the basis for calculating Kendall's τ coefficient. Simulations are performed with a wide variety of real complex networks in different dimensions. In addition to traditional centrality metrics, we use several state-of-the-art centrality metrics to compare with WSLC. The rest of this section includes details related to experimental setup, evaluation criteria, benchmark metrics and simulation results.

4.1. Experimental setup

All simulations were performed on a Lenovo laptop with Intel® Core™ i5 Processor (2.40 GHz), 8 GB of Memory, and Windows 11 Home. MATLAB R2023a software is used to implement centrality metrics. Each centrality metric is tested on eight real complex networks, where we select *TopK* nodes with the highest rank from each network as influential nodes. Inspired by Lv et al. (2019), we set *TopK* equal to 10 in the simulations.

Algorithm 2. Pseudocode of WSLC centrality metric

Input: $G = (V, E)$, L, β, α_1, α_2, α_3, node v.
Output: $WSLC(v)$ as the influence of node v.
1: $N = |V|$, $M = |E|$;
2: $A \leftarrow$ Adjacency matrix of graph G;
3: $\{w_{i,j}\}_{N\times N} \leftarrow$ Adjusting the weight between each pair of nodes i and j according to a weighted policy;
// Definition of Node-Influence
4: $D_G = 2M/(N(N-1))$;
5: $I_{Node}(v) = k_v/(k_{max} + D_G)$;
// Definition of Local-Influence
6: $\Gamma(v) \leftarrow$ Find the set of all first-level neighbors of v based on A;
7: $I_{Local}(v) = 0$;
8: **foreach** $u \in \Gamma(v)$ **do**
9: $I_{Local}(v) = I_{Local}(v) + \sqrt{w_{u,v} {}^* k_v}/(k_u + k_v)$;
10: **end**
11: $I_{Local}(v) = I_{Local}(v)/k_v$;
// Definition of Semi-Local-Influence
12: $I_{Semi\_Local}(v) = 0$;
13: $G_v^{<L>} \leftarrow$ Extracting the semi-local subgraph associated with v at level l according to extended neighborhood concept;
14: **for** $l = 2$ to L **do**
15: $\Gamma_l(v) \leftarrow$ Find the set of all neighbors up to the maximum level l of v according to extended neighborhood concept;
16: **foreach** $u \in \Gamma_l(v)$ **do**
17: $Path_{u,v} \leftarrow shortestpath(G_v^{<L>}, u, v)$;
18: $\widehat{w}_{u,v} = 1$; $tmp = 0$;
19: **foreach** $e_{i,j} \in Path_{u,v}$ **do**
20: $\widehat{w}_{u,v} = \widehat{w}_{u,v} {}^* w_{i,j}$;
21: **end**
22: $\delta_{u,v} = |Path_{u,v}|$;
23: $tmp = tmp + \sqrt{\widehat{w}_{u,v} {}^* k_v}/(\delta_{u,v} {}^* (k_u + k_v))$;
24: **end**
25: $I_{Semi\_Local}(v) = I_{Semi\_Local}(v) + (\beta^l {}^* tmp)$
26: **end**
27: $I_{Semi\_Local}(v) = I_{Semi_{Local}}(v)/SZ[G_v^{<L>}]$;
// Definition of Total-Influence
28: $WSLC(v) = \alpha_1 {}^* I_{Node}(v) + \alpha_2 {}^* I_{Local}(v) + \alpha_2 {}^* I_{Semi\_Local}(v)$;
29: **Return** $WSLC(v)$;

In addition to the edge weight policy, α_1, α_2, α_3, and L are tunable parameters in the proposed metric. We examine several weighting policies to find the best performance. Also, we analyze the parameter L through simulation with different values. Parameters α_1, α_2, and α_3, which are applied as influence coefficients in the proposed metric, are set in the optimal mode using Taguchi design approach (Liu et al., 2022). This approach seeks to maximize the S/N ratio (signal-to-noise) by considering the influence score. We found the optimal values for parameters α_1, α_2, and α_3 with Taguchi approach 0.25, 0.3 and 0.45, respectively.

Table 5
Properties of the real complex networks.

Networks	No. of nodes	No. of edges	Avg. degree	Max. degree
Karate-Club	34	78	4.6	17
Dolphins	62	159	5.1	12
C-elegans	297	2148	15.2	83
Airlines	235	1297	11.0	130
Infect-Dub	410	2765	13.5	50
Email	1133	5451	9.62	71
Grid	4941	6594	2.7	19
Ca-Astroph	18,771	198,050	22.0	504

4.2. Description of complex networks

In total, eight real complex networks are used in the experiments: Karate-Club, Dolphins, C-elegans, Airlines, Infect-Dub, Email, Grid, and Ca-Astroph. We assume all these networks are undirected and unweighted. Also, these networks are selected based on different sizes to evaluate the performance of centrality metrics in different conditions. Topological properties of all selected networks are available via https://networkrepository.com/networks.php. Table 5 provides details about the topological properties of these networks.

4.3. Evaluation criteria

Models such as Susceptible-Infected-Recovered (SIR) are widely used to investigate the way of spreading information in the entire network by nodes with high ranking (Liu et al., 2016). Each centrality metric selects a number of nodes that have the highest influence score as top nodes. These nodes are considered as initial infected nodes in the SIR model. The process of spreading information is applied by the SIR model according to the probability of infection λ and the probability of recovery μ as well as initial infected nodes. Here, each infected node can infect its susceptible neighbors at rate λ. Meanwhile, infected nodes can recover at rate μ. The process of spreading information in SIR is performed at each step t until there are no infected nodes in the network.

At the beginning of the simulation, all nodes except initial infected nodes are selected as susceptible nodes. At each time step t, each infected node randomly selects only one of its susceptible neighbors and then infects it with probability λ. According to Zhang et al. (2023), we analyze the probability of infection λ in the range of 0.01 to 0.1 while μ is set to 1. Let $F(t)$ be the set of infected and recovered nodes at step t. As shown in Eq. (12), $F(t)$ can be considered as an index to evaluate the influence of nodes in step t.

$$F(t) = \frac{N_{I(t)} + N_{R(t)}}{N} \tag{12}$$

where $N_{I(t)}$ is the number of infected nodes and $N_{R(t)}$ is the number of recovered nodes.

Kendall's τ coefficient has been used as a measure for correlation analysis of ranking lists in extensive studies (Liu et al., 2016). In the problem of identifying influential nodes, Kendall's τ coefficient can be applied to evaluate the correlation between different centrality metrics. Let $R = \{r_1, r_2, \cdots, r_N\}$ be the ranking list provided for N nodes by a centrality metric. Also, let $R' = \{r'_1, r'_2, \cdots, r'_N\}$ be the ranking list generated for N nodes by the SIR model. If $(r_i < r_j)$ and $(r'_i < r'_j)$ or $(r_i > r_j)$ and $(r'_i > r'_j)$ then pair $(r_i < r'_i)$ and $(r_j < r'_j)$ said to be concordant. Also, if $(r_i < r_j)$ and $(r'_i > r'_j)$ or $(r_i > r_j)$ and $(r'_i < r'_j)$ then pair $(r_i < r'_i)$ and $(r_j < r'_j)$ said to be discordant. In Kendall's coefficient, τ is defined by Eq. (13), where its higher value indicates the similar behavior of two lists R and R'.

$$\tau = \frac{N_c - N_d}{0.5N(N-1)} \tag{13}$$

where N_c is the number of concordant pairs and N_d is the number of discordant pairs.

4.4. Benchmark metrics

We use different centrality metrics from all three categories of local, semi-local, and global to compare with the proposed metric. In the local category, we compare WSLC with DC (Freeman, 2002), PR (Brin and Page, 1998), and TPR (Sheng et al., 2020b). In the global category, we use BC (Freeman, 1977), CC (Sabidussi, 1966), EC (Bonacich, 2007), and RASP (Lv et al., 2019) to validate the proposed centrality metric. Also, several state-of-the-art and equivalent centrality metrics from the semi-local category such as SC (Chen et al., 2012), WSLCC (Kang et al., 2016), LGC (Ullah et al., 2021), and INASP (Zhang et al., 2023) are used to compare with WSLC. For fair comparisons, we implement all available centrality metrics by the same experimental setup.

4.5. Simulation results

In this section, we prove with numerical simulations and various comparisons that WSLC performs much better than traditional and state-of-the-art centrality metrics.

The most important tunable parameter in WSLC is L, which determines the nearest neighbor level. Here, $L = 1$ indicates that WSLC uses only first-level neighbors to calculate the influence of each node in the network. Setting L to 2 means that the information of all first and second level nodes are considered for ranking. Likewise, an increase in L will lead to an increase in the level of neighborhood and an increase in the information available to measure influence. However, configuring WSLC with high neighborhood levels leads to increased complexity, as information for more nodes must be processed. On the other hand, the information of nodes with higher neighborhood levels has less value in ranking. It is obvious that using too high neighborhood levels leads to the addition of useless information and thus to the performance reduction. Hence, setting the parameter L plays an important role in identifying the influential nodes. In a comparative experiment, we examined values of L from 1 to 5. The results of this comparison for all complex networks are reported in Fig. 2. The average results show that WSLC with a neighborhood level of 4 has the best performance. As illustrated, $L = 2$ has led to much better results than $L = 1$. Meanwhile, increasing L up to 4 improves the WSLC results, while the performance of WSLC is degraded for $L > 4$. These results confirm the effectiveness of the extended neighborhood concept for identifying neighboring nodes, because considering neighbors with different and limited levels has led to improved results. This simulation is done with $\lambda = 0.1$, $\mu = 1$, and $w_{u,v} = ND_{u,v}$.

Another important parameter in WSLC is the edge weighting policy, which applies the contribution of connections between nodes to the ranking. We argued that considering the equal contribution of edges in measuring influence is an unrealistic assumption of social interactions. Hence, we prove this claim by a numerical experiment. In this experiment, we apply several different weights as an edge weighting policy on WSLC. Also, we set the WSLC for each pair of nodes u and v with $w_{u,v} = 1$, which leads to applying the same contribution of edges in the node ranking process. The results of this experiment can show the effect of weighting the edges as well as the policy of choosing the optimal weight.

Table 6 shows the results related to Kendall's τ coefficient for six defined weighting policies. Each row shows the results for a complex network and the last row is dedicated to the average results. Each column is a weighting policy in WSLC, and the column associated with '$w_{u,v} = 1$' is the results of the unweighted version of WSLC. This simulation is done with $L = 4$, $\lambda = 0.1$, and $\mu = 1$. As depicted, the WSLC setting with $w_{u,v} = 1$ provides little performance in ranking, and this confirms our claim about the different contribution of edges in measuring influence. Meanwhile, the results clearly show that $ND_{u,v}$

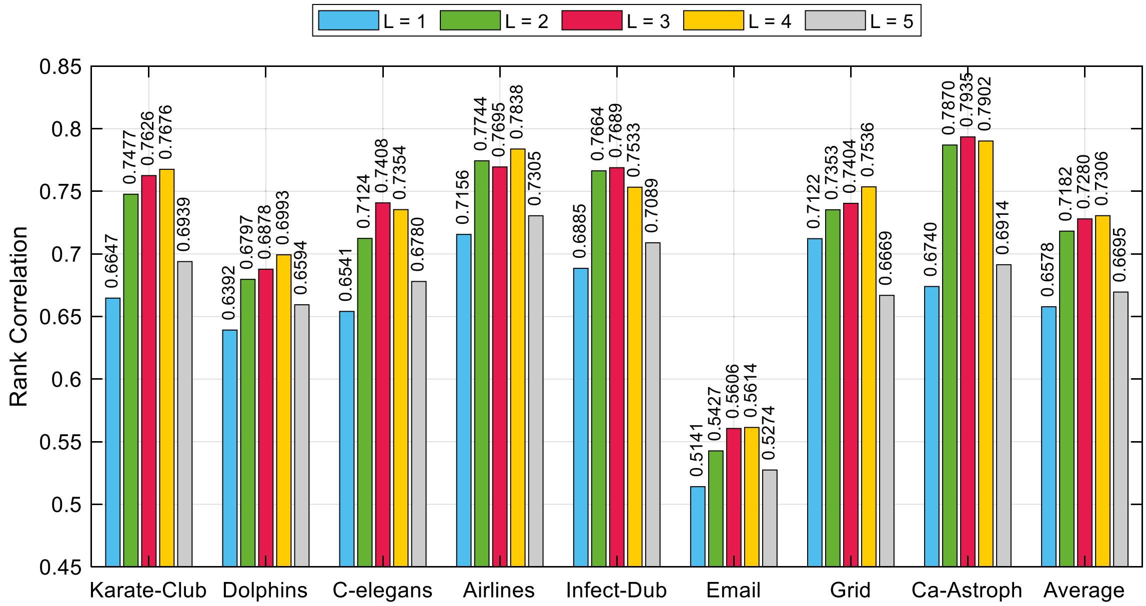

Fig. 2. Analysis of the parameter L in the proposed WSLC metric.

Table 6
Performance of WSLC with different edge weighting policies.

Networks	$w_{u,v} = 1$	$CN_{u,v}$	$JC_{u,v}$	$AD_{u,v}$	$ND_{u,v}$	$RO_{u,v}$	$KI_{u,v}$
Karate-Club	0.7649	0.7756	0.7695	0.7592	0.7676	0.7416	0.7701
Dolphins	0.6829	0.6971	0.6967	0.6871	0.6993	0.6789	0.6941
C-elegans	0.6796	0.7018	0.7015	0.7166	0.7354	0.7008	0.6964
Airlines	0.775	0.7888	0.7877	0.7826	0.7838	0.7817	0.7863
Infect-Dub	0.7392	0.7553	0.7583	0.7458	0.7533	0.7368	0.7511
Email	0.5565	0.5632	0.5641	0.5602	0.5614	0.5524	0.5587
Grid	0.7323	0.7556	0.7595	0.7394	0.7536	0.7540	0.7558
Ca-Astroph	0.7732	0.7892	0.7927	0.7795	0.7902	0.7687	0.7873
Average	0.7130	0.7283	0.7288	0.7213	0.7306	0.7144	0.7250

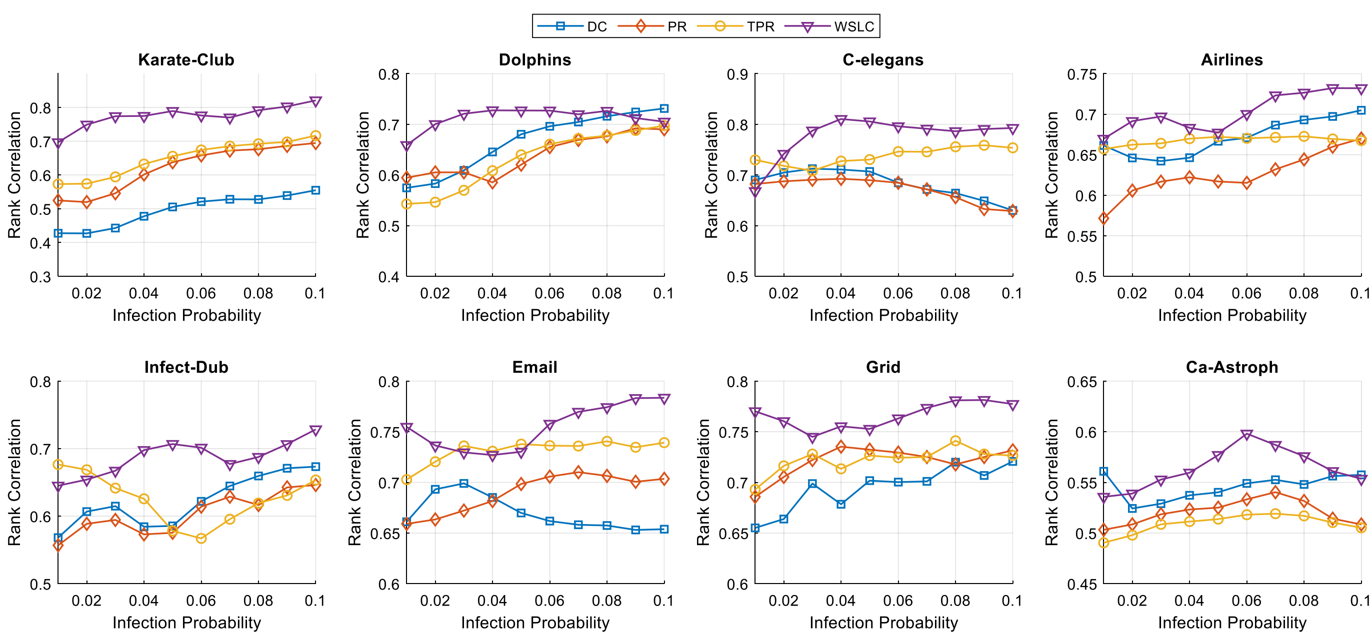

Fig. 3. Comparison of WSLC with local centrality metrics (i.e., DC, PR and TPR) based on Kendall's τ coefficient.

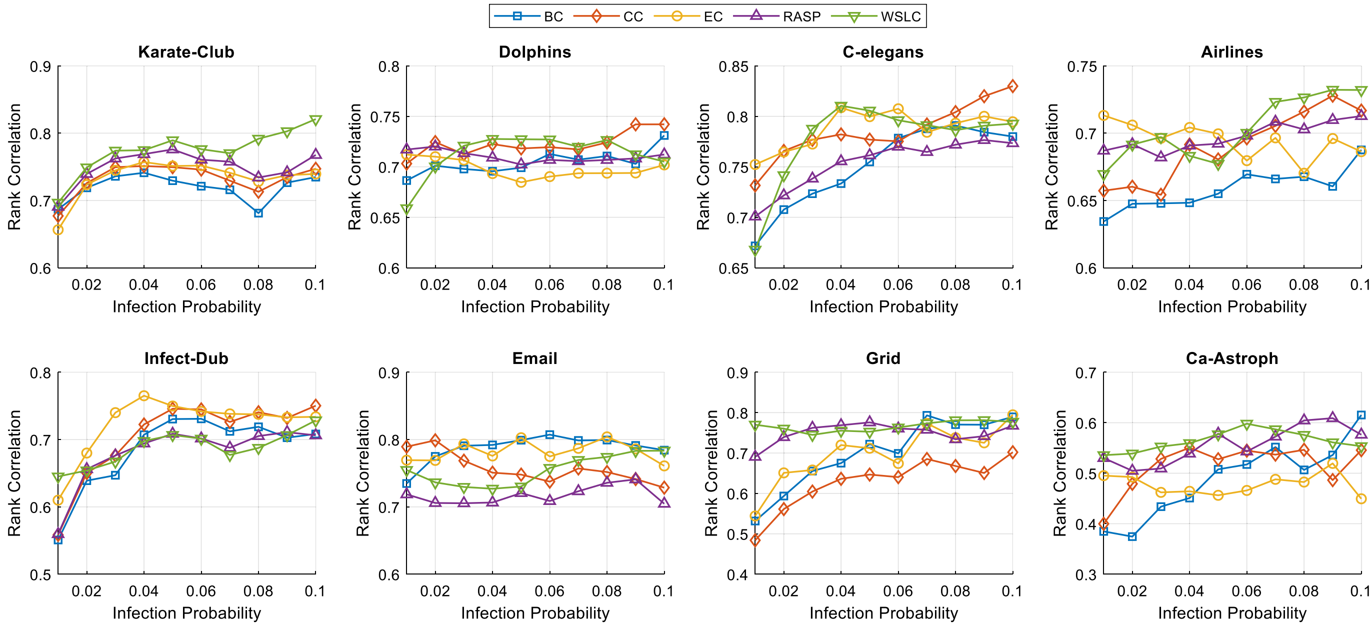

Fig. 4. Comparison of WSLC with global centrality metrics (i.e., BC, CC, EC and RASP) based on Kendall's τ coefficient.

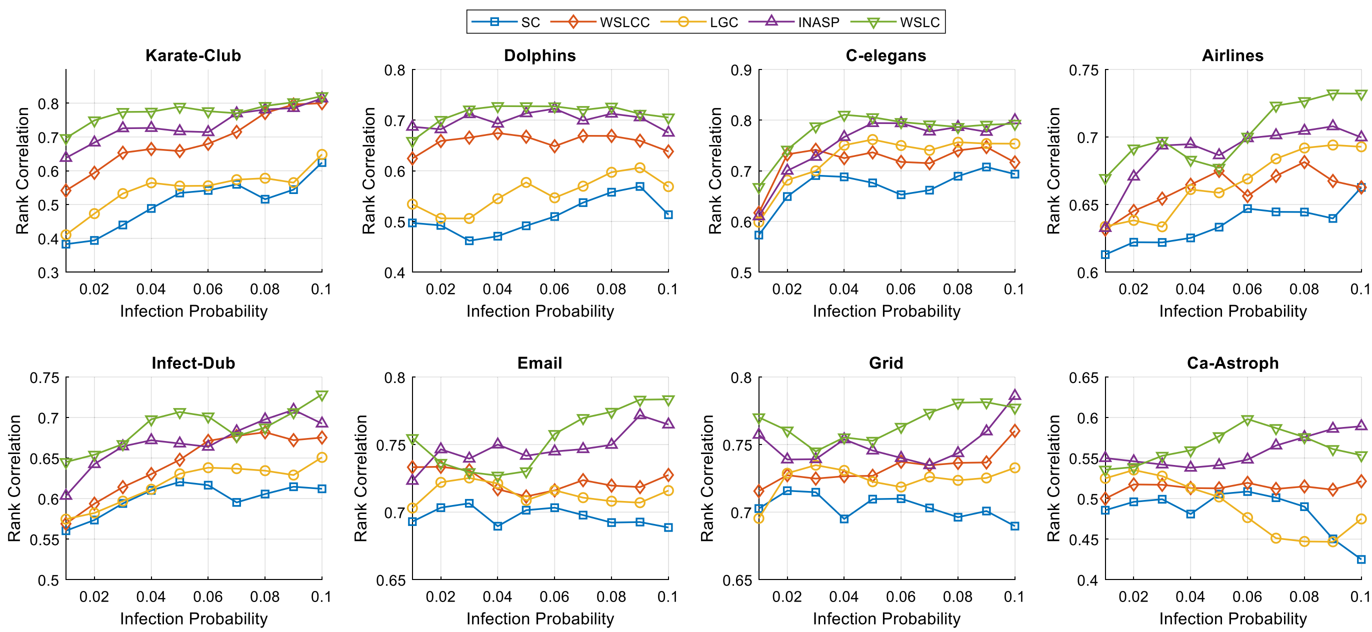

Fig. 5. Comparison of WSLC with semi-local centrality metrics (i.e., SC, WSLCC, LGC, and INASP) based on Kendall's τ coefficient.

provides the best weighting policy for WSLC.

The performance of WSLC compared to local centrality metrics (i.e., DC, PR and TPR) is presented in Fig. 3 for all complex networks. In this comparison, Kendall's τ coefficient is reported for each metric based on different rates of λ (0.01 to 0.1). Each probability of λ expresses the relationship between centrality metrics and $F(t)$. In this figure, rank correlation represents τ in Kendall's coefficient, where it depicts the correlation associated with cumulative infected nodes for centrality metrics. Comparison of WSLC with global centrality metrics (i.e., BC, CC, EC and RASP) based on Kendall's τ coefficient is given in Fig. 4. Also, Fig. 5 evaluates the performance of WSLC against semi-local centrality metrics (i.e., SC, WSLCC, LGC, and INASP). For better clarity of comparisons, the average Kendall's τ coefficient of each metric was calculated on all complex networks and reported in Fig. 6. Fig. 6(a) shows the average results of Kendall's τ coefficient for WSLC as well as DC, PR and TPR metrics. Fig. 6(b) shows the average results for WSLC as well as other global metrics such as BC, CC, EC and RASP. In addition, Fig. 6(c) shows the average results related to Kendall's τ coefficient for SC, WSLCC, LGC, INASP and WSLC.

As illustrated, WSLC significantly identifies influential nodes more accurately compared to local centrality metrics, as it results in $F(t)$ with higher correlation than other metrics. Actually, WSLC as a semi-local metric uses more information than local metrics and its higher performance is expected. However, WSLC produces competitive results in almost all cases compared to global centrality metrics. As we can observe, the average results of WSLC are not decisively superior compared to other global metrics, because global metrics use all network information to measure influence. However, these metrics have significantly higher time complexity. On the other hand, in some cases, the use of all network information may lead to weakening the performance

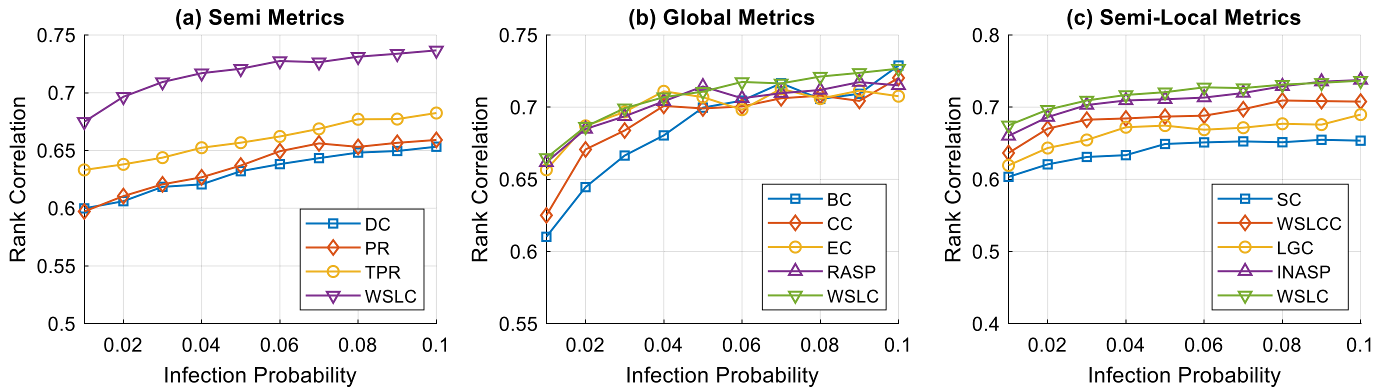

Fig. 6. Average results of Kendall's τ coefficient on all complex networks. (a) Comparison of WSLC with local metrics, (b) Comparison of WSLC with global metrics, and (c) Comparison of WSLC with semi-local metrics.

of a centrality metric. For example, BC and CC perform worse than the proposed metric in some networks. In addition, WSLC provides better results as networks scale up compared to other metrics, because WSLC only considers a subgraph of the network for ranking nodes. Therefore, increasing the scale of the network does not have a significant effect on the results provided by WSLC. It is worth noting that WSLC uses the extended neighborhood concept based on a distributed approach to extract subgraphs, which imposes little complexity on the ranking process. As we can observe in most of the networks, WSLC as a semi-local metric provides better results compared to its equivalent metrics. The reason for this superiority is the use of neighborhood information with different levels and also considering the importance of both nodes and edges. On average, WSLC outperforms SC, WSLCC, LGC, and INASP by 12.8 %, 4.6 %, 7.3 %, and 2.2 %, respectively.

Time complexity is one of the effective factors in evaluating centrality metrics. The runtime results as one of the time complexity indicators for the centrality metrics are given in Table 7. The results of this simulation are reported based on $w_{u,v} = ND_{u,v}$, $L = 4$, $\lambda = 0.1$, and $\mu = 1$. These results include the running time to identify 10 influential nodes, where we do not consider the loading time of the dataset in the simulation. DC has the lowest running time compared to other metrics with an average of 0.12 s. DC only needs the degree information of the nodes and is a local metric. Since the degree of the nodes is available through the adjacency matrix, DC does not depend on the network size and has very low running time. Similarly, PR and TPR are based only on the information of first-level neighbors and therefore have a short runtime. Instead, BC, CC, EC, and RASP as global metrics have a very high runtime. Since these metrics need to process the information of the entire network to measure the influence of a node, their running time increases with the increase of the network size.

Other metrics available as semi-local metrics have different runtimes. LGC and WSLCC have similar performance in runtime results because they use practically the same information to calculate centrality. Meanwhile, the average runtime results in SC are also similar. However, the running time in SC has little change with increasing network size, because this metric only uses first and second level neighbors. The runtime results in SC, INASP and WSLC are highly competitive. SC requires only the degree of first and second level neighbors and is therefore less complex than other metrics. WSLC is less complex than INASP because it uses a distributed approach to extract local subgraphs. On average, WSLC requires 4.1 s and 15.6 s less runtime compared to SC and INASP, respectively.

5. Conclusions

The purpose of this paper is to identify influential nodes in complex networks and understand the theoretical and practical importance of centrality metrics. So far, various centrality metrics have been proposed to solve this problem, so that the performance of each depends on specific scenarios. For example, metrics based on local structure have low ranking accuracy due to the use of limited information, and metrics based on global structure suffer from high complexity. Meanwhile, metrics based on semi-local structure are amazingly well, but an efficient centrality for identifying influential nodes is still not available due to differences in the structure and scale of networks. In addition, most semi-local centrality metrics only consider one aspect of each node's information, and their development still faces serious challenges. It is obvious that the importance of the edge and its stability play a significant role in the ranking of nodes. In addition, the effectiveness of the extended neighborhood concept has been confirmed by many researchers to consider different neighborhood levels in the influence calculation. Considering all these issues, we proposed a weighted semi-local centrality metric called WSLC to improve the identification of influential nodes in complex networks. WSLC considers the importance of both nodes and edges simultaneously to calculate influence. The importance of edge is covered by studying different number of weights.

Table 7
Runtime results (seconds) for WSLC and other centrality metrics.

Centrality metrics	Metric type	Karate-Club	Dolphins	C-elegans	Airlines	Infect-Dub	Email	Grid	Ca-Astroph	Average
DC	Local	0.02	0.06	0.07	0.11	0.12	0.14	0.16	0.26	0.12
PR	Local	0.46	0.84	1.10	1.26	1.33	1.65	1.88	3.44	1.50
TPR	Local	0.70	1.03	1.43	1.50	1.67	2.00	2.16	4.26	1.84
BC	Global	3.05	14.15	17.52	17.86	19.05	24.07	77.34	154.2	40.91
CC	Global	2.98	13.69	17.26	18.25	18.67	23.68	80.13	155.7	41.30
EC	Global	2.30	8.75	13.04	14.11	15.26	18.85	66.00	118.5	32.10
RASP	Global	2.84	9.15	11.76	13.48	15.56	21.34	55.72	124.3	31.77
SC	Semi-local	0.87	3.30	3.84	3.87	4.23	7.12	21.46	34.82	9.94
LGC	Semi-local	2.65	5.38	6.03	6.35	6.99	10.33	48.27	77.08	20.39
WSLCC	Semi-local	3.34	6.64	9.00	10.41	11.37	15.68	44.13	76.21	22.10
INASP	Semi-local	1.26	5.11	5.97	5.68	6.06	8.21	22.16	33.83	11.04
WSLC	Semi-local	0.66	2.73	4.05	5.45	5.74	6.32	20.30	31.17	9.55

Also, WSLC uses the extended neighborhood concept to be efficient when dealing with large-scale networks. Here, a subgraph of the network is extracted with a distributed approach to determine nearest neighbors with different hops. According to the considered information, WSLC combines the topological position with the semi-local structure to perform better to identify the influential nodes.

A comparative analysis of WSLC with several classical and equivalent centrality metrics has been performed on eight real-world complex networks. The insight of using the extended neighborhood concept to determine nearest neighbors has resulted in higher accuracy and lower complexity simultaneously. WSLC reports better results in terms of Kendall's correlation coefficient and is more efficient in dealing with large-scale networks. We compared the cumulative infected nodes obtained from the SIR model with WSLC and other state-of-the-art metrics. The simulation results showed that the proposed WSLC metric is not correlated with the existing centrality metrics. Also, our metric provides better results compared to DC, PR, TPR, BC, CC, EC, RASP, SC, WSLCC, LGC, and INASP. To summarize, the reasons for the superiority of the proposed method can be listed as follows: 1) Simultaneously considering the importance of the node itself and its nearest neighbors to calculate the influence; 2) Identifying nearest neighbors with a low-complexity distributed approach; 3) Using nearest neighbors with different levels to rank nodes; 4) Using a damping coefficient to apply a higher effect of closer neighbors; 5) Applying the importance of each edge in the calculation of influence by assigning weight to the edges.

In addition to theoretical significance, the identification of influential nodes also has practical applications. Considering the acceptable performance of the proposed metric for identifying influential nodes in complex networks, WSLC can be considered for real-world networks. For example, WSLC can be used for optimal ranking of search engine results. Influential nodes selected by WSLC can be used as centers to detect communities in social networks. The ranking of nodes by WSLC can be used for the purpose of targeted backup according to the importance of server nodes in computer networks, where this leads to ensuring the robustness of the network. To improve the proposed metric, there are some potential issues that can be considered as future work. For example, other features such as interactions between nodes can be considered as edge weights. Also, WSLC can be extended and adapted to dynamic and directed networks. Developing parallel approaches to identify influential nodes is another future direction. Considering the relative change in average shortest path theory with extended neighborhood concept is clearly neglected in the existing literature.

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.

References

Berahmand, K., Bouyer, A., Vasighi, M., 2018. Community detection in complex networks by detecting and expanding core nodes through extended local similarity of nodes. IEEE Trans. Comput. Social Syst. 5 (4), 1021–1033.

Berahmand, K., Nasiri, E., Forouzandeh, S., Li, Y., 2022. A preference random walk algorithm for link prediction through mutual influence nodes in complex networks. J. King Saud Univ.-Comput. Inf. Sci. 34 (8), 5375–5387.

Bonacich, P., 2007. Some unique properties of eigenvector centrality. Soc. Networks 29 (4), 555–564.

Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual web search engine. Comput. Networks ISDN Syst. 30 (1–7), 107–117.

Cao, C., Wang, J., Kwok, D., Cui, F., Zhang, Z., Zhao, D., et al., 2022. webTWAS: a resource for disease candidate susceptibility genes identified by transcriptome-wide association study, Nucl. Acids Res. 50(D1), D1123-D1130.

Cao, Y., Xu, N., Wang, H., Zhao, X., Ahmad, A.M., 2023. Neural networks-based adaptive tracking control for full-state constrained switched nonlinear systems with periodic disturbances and actuator saturation. Int. J. Syst. Sci. 54 (14), 2689–2704.

Chen, D., Lü, L., Shang, M.S., Zhang, Y.C., Zhou, T., 2012. Identifying influential nodes in complex networks. Physica A 391 (4), 1777–1787.

Dai, J., Wang, B., Sheng, J., Sun, Z., Khawaja, F.R., Ullah, A., Duan, G., 2019. Identifying influential nodes in complex networks based on local neighbor contribution. IEEE Access 7, 131719–131731.

Fan, C., Zeng, L., Sun, Y., Liu, Y.Y., 2020. Finding key players in complex networks through deep reinforcement learning. Nat. Mach. Intell. 2 (6), 317–324.

Forouzandeh, S., Rostami, M., Berahmand, K., 2021. Presentation a Trust Walker for rating prediction in recommender system with Biased Random Walk: effects of H-index centrality, similarity in items and friends. Eng. Appl. Artif. Intel. 104, 104325.

Freeman, L.C., 1977. A set of measures of centrality based on betweenness. Sociometry 35–41.

Freeman, L.C., 2002. Centrality in social networks: conceptual clarification. In: Social network: critical concepts in sociology. Routledge, Londres, pp. 238–263.

Gao, S., Ma, J., Chen, Z., Wang, G., Xing, C., 2014. Ranking the spreading ability of nodes in complex networks based on local structure. Physica A 403, 130–147.

Gao, C., Wei, D., Hu, Y., Mahadevan, S., Deng, Y., 2013. A modified evidential methodology of identifying influential nodes in weighted networks. Physica A 392 (21), 5490–5500.

Guo, S., Zhao, X., Wang, H., Xu, N., 2023. Distributed consensus of heterogeneous switched nonlinear multiagent systems with input quantization and DoS attacks. Appl. Math Comput. 456, 128127.

Hajarathaiah, K., Enduri, M.K., Anamalamudi, S., 2022. Efficient algorithm for finding the influential nodes using local relative change of average shortest path. Physica A 591, 126708.

Huang, S., Zong, G., Wang, H., Zhao, X., Alharbi, K.H., 2023. Command filter-based adaptive fuzzy self-triggered control for MIMO nonlinear systems with time-varying full-state constraints. Int. J. Fuzzy Syst. https://doi.org/10.1007/s40815-023-01560-8.

Jaccard, P., 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bull. Soc. Vaudoise Sci. Nat. 37, 547–579.

Jannesari, V., Keshvari, M., Berahmand, K., 2023. A novel nonnegative matrix factorization-based model for attributed graph clustering by incorporating complementary information. Expert Syst. Appl. 242, 122799.

Kang, W., Tang, G., Sun, Y., Wang, S., 2016. Identifying influential nodes in complex network based on weighted semi-local centrality. In: 2016 2nd IEEE International Conference on Computer and Communications (ICCC). IEEE, pp. 2467–2471.

Katz, L., 1953. A new status index derived from sociometric analysis. Psychometrika 18 (1), 39–43.

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

Kumar, S., Panda, B.S., 2020. Identifying influential nodes in Social Networks: neighborhood coreness based voting approach. Physica A 553, 124215.

Li, Z., Ren, T., Ma, X., Liu, S., Zhang, Y., Zhou, T., 2019. Identifying influential spreaders by gravity model. Sci. Rep. 9 (1), 8387.

Liu, C., Wang, J., Zhou, L., Rezaeipanah, A., 2022. Solving the multi-objective problem of IoT service placement in fog computing using cuckoo search algorithm. Neural Process. Lett. 54 (3), 1823–1854.

Liu, J., Xiong, Q., Shi, W., Shi, X., Wang, K., 2016. Evaluating the importance of nodes in complex networks. Physica A 452, 209–219.

Lorrain, F., White, H.C., 1971. Structural equivalence of individuals in social networks. J. Math. Sociol. 1 (1), 49–80.

Lv, Z., Zhao, N., Xiong, F., Chen, N., 2019. A novel measure of identifying influential nodes in complex networks. Physica A 523, 488–497.

Masdari, M., Gharehpasha, S., Ghobaei-Arani, M., Ghasemi, V., 2020. Bio-inspired virtual machine placement schemes in cloud computing environment: taxonomy, review, and future research directions. Clust. Comput. 23 (4), 2533–2563.

Mohammadian, E., Dastgerdi, M.E., Manshad, A.K., Mohammadi, A.H., Liu, B., Iglauer, S., Keshavarz, A., 2022. Application of underbalanced tubing conveyed perforation in horizontal wells: a case study of perforation optimization in a giant oil field in Southwest Iran. Adv. Geo-Energy Res. 6, 296–305.

Ni, Q., Guo, J., Wu, W., Wang, H., Wu, J., 2021. Continuous influence-based community partition for social networks. IEEE Trans. Network Sci. Eng. 9 (3), 1187–1197.

Rezaeipanah, A., Ahmadi, G., Sechin Matoori, S., 2020. A classification approach to link prediction in multiplex online ego-social networks. Soc. Netw. Anal. Min. 10 (1), 27.

Rostami, M., Oussalah, M., Berahmand, K., Farrahi, V., 2023. Community detection algorithms in healthcare applications: a systematic review. IEEE Access 11, 30247–30272.

Sabidussi, G., 1966. The centrality index of a graph. Psychometrika 31 (4), 581–603.

Salavati, C., Abdollahpouri, A., Manbari, Z., 2019. Ranking nodes in complex networks based on local structure and improving closeness centrality. Neurocomputing 336, 36–45.

Serrano, M.Á., Boguna, M., 2006. Clustering in complex networks. I. General formalism. Phys. Rev. E 74 (5), 056114.

Shao, Z., Liu, S., Zhao, Y., Liu, Y., 2019. Identifying influential nodes in complex networks based on Neighbours and edges. Peer-to-Peer Netw. Appl. 12, 1528–1537.

Sheng, J., Dai, J., Wang, B., Duan, G., Long, J., Zhang, J., Guan, W., 2020a. Identifying influential nodes in complex networks based on global and local structure. Physica A 541, 123262.

Sheng, J., Zhu, J., Wang, Y., Wang, B., Hou, Z.A., 2020b. Identifying influential nodes of complex networks based on trust-value. Algorithms 13 (11), 280.

Shetty, R.D., Bhattacharjee, S., 2022. A Weighted Hybrid Centrality for Identifying Influential Individuals in Contact Networks. In: 2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT). IEEE, pp. 1–6.

Torabi, E., Ghobaei-Arani, M., Shahidinejad, A., 2022. Data replica placement approaches in fog computing: a review. Clust. Comput. 25 (5), 3561–3589.

Ullah, A., Wang, B., Sheng, J., Long, J., Khan, N., Sun, Z., 2021. Identifying vital nodes from local and global perspectives in complex networks. Expert Syst. Appl. 186, 115778.
Ullah, A., Wang, B., Sheng, J., Khan, N., 2022. Escape velocity centrality: escape influence-based key nodes identification in complex networks. Appl. Intell. 52 (14), 16586–16604.
Wang, S., Sheng, H., Yang, D., Zhang, Y., Wu, Y., Wang, S., 2022. Extendable multiple nodes recurrent tracking framework with RTU++. IEEE Trans. Image Process. 31, 5257–5271.
Wang, T., Zhang, L., Xu, N., Alharbi, K.H., 2023. Adaptive critic learning for approximate optimal event-triggered tracking control of nonlinear systems with prescribed performances. Int. J. Control. https://doi.org/10.1080/00207179.2023.2250880.
Wu, W., Xu, N., Niu, B., Zhao, X., Ahmad, A.M., 2023. Low-computation adaptive saturated self-triggered tracking control of uncertain networked systems. Electronics 12 (13), 2771.
Xue, B., Yang, Q., Jin, Y., Zhu, Q., Lan, J., Lin, Y., Tan, J., Liu, L., Zhang, T., Chirwa, E.M. N., Zhou, X., 2023. Genotoxicity Assessment of Haloacetaldehyde Disinfection Byproducts via a Simplified Yeast-Based Toxicogenomics Assay. Environ. Sci. Technol. 57 (44), 16823–16833.
Yang, R., Jia, A., Hu, Q., Guo, X., Sun, M., 2020. Particle size effect on water vapor sorption measurement of organic shale: one example from Dongyuemiao Member of Lower Jurassic Ziliujing Formation in Jiannan Area of China. Adv. Geo-Energy Res. 4 (2), 207–218.
Yang, R., Yang, C., Peng, X., Rezaeipanah, A., 2022. A novel similarity measure of link prediction in multi-layer social networks based on reliable paths. Concurrency Comput.: Practice Exp. 34 (10), e6829.
Yue, S., Niu, B., Wang, H., Zhang, L., Ahmad, A.M., 2023. Hierarchical sliding mode-based adaptive fuzzy control for uncertain switched under-actuated nonlinear systems with input saturation and dead-zone. Robotic Intell. Autom. 43 (5), 523–536.
Zeng, A., Zhang, C.J., 2013. Ranking spreaders by decomposing complex networks. Phys. Lett. A 377 (14), 1031–1035.
Zhang, K., Zhou, Y., Long, H., Wang, C., Hong, H., Armaghan, S.M., 2023. Towards identifying influential nodes in complex networks using semi-local centrality metrics. J. King Saud Univ.-Comput. Inf. Sci. 35 (10), 101798.
Zhang, H., Zou, Q., Ju, Y., Song, C., Chen, D., 2022. Distance-based support vector machine to predict DNA N6-methyladenine modification. Curr. Bioinform. 17 (5), 473–482.
Zhao, Y., Liang, H., Zong, G., Wang, H., 2023c. Event-Based Distributed Finite-Horizon H∞Consensus Control for Constrained Nonlinear Multiagent Systems. IEEE Syst. J. 17 (4), 5369–5380.
Zhao, Y., Niu, B., Zong, G., Zhao, X., Alharbi, K.H., 2023b. Neural network-based adaptive optimal containment control for non-affine nonlinear multi-agent systems within an identifier-actor-critic framework. J. Franklin Inst. 360 (12), 8118–8143.
Zhao, J., Wang, Y., Deng, Y., 2020. Identifying influential nodes in complex networks from global perspective. Chaos Solitons Fractals 133, 109637.
Zhao, H., Wang, H., Xu, N., Zhao, X., Sharaf, S., 2023a. Fuzzy approximation-based optimal consensus control for nonlinear multiagent systems via adaptive dynamic programming. Neurocomputing 553, 126529.
Zhao, J., Wen, T., Jahanshahi, H., Cheong, K.H., 2022. The random walk-based gravity model to identify influential nodes in complex networks. Inf. Sci. 609, 1706–1720.
Zhao, X., Xing, S., Wang, Q., 2018. Identifying influential spreaders in social networks via normalized local structure attributes. IEEE Access 6, 66095–66104.
Zhao, H., Zong, G., Wang, H., Zhao, X., Xu, N., 2023d. Zero-Sum Game-Based Hierarchical Sliding-Mode Fault-Tolerant Tracking Control for Interconnected Nonlinear Systems via Adaptive Critic Design. IEEE Trans. Autom. Sci. Eng. https://doi.org/10.1109/TASE.2023.3317902.
Zhao, H., Zong, G., Zhao, X., Wang, H., Xu, N., Zhao, N., 2023e. Hierarchical Sliding-Mode Surface-Based Adaptive Critic Tracking Control for Nonlinear Multiplayer Zero-Sum Games Via Generalized Fuzzy Hyperbolic Models. IEEE Trans. Fuzzy Syst. 31 (11), 4010–4023.
Zhou, G., Deng, R., Zhou, X., Long, S., Li, W., Lin, G., Li, X., 2021. Gaussian inflection point selection for LiDAR hidden echo signal decomposition. IEEE Geosci. Remote Sens. Lett. 19, 1–5.